A Novel Fast Packet Switch Architecture For ATM Networks

by

Wasif Hasan

A Thesis Presented to the

FACULTY OF THE COLLEGE OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

COMPUTER ENGINEERING

June, 1996

A Novel Fast Packet Switch Architecture For ATM Networks

BY

Wasif Hasan

A Thesis Presented to the

FACULTY OF THE COLLEGE OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE
In
Computer Engineering

June, 1996.

A Novel Fast Packet Switch Architecture For ATM Networks

Wasif Hasan

Computer Engineering

June, 1996

KING FAHD UNIVERSITY OF PETROLEUM AND MINERALS

DHAHRAN, SAUDI ARABIA

COLLEGE OF GRADUATE STUDIES

This thesis, written by

Wasif Hasan

under the direction of his Thesis Advisor, and approved by his Thesis committee, has been presented to and accepted by the Dean, College of Graduate Studies, in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN COMPUTER ENGINEERING

Thesis Committee :

Dr. Mayez Al-Mouhamed (Chairman)

Dr. Habib Youssef (Co-Chairman)

for Dr. Muhammad S. T. Benten
Department Chairman (Comp. Engg.)

Dr. Mostafa Abd-El-Barr (Member)

Dr. Ala H. Al-Rabeh
Dean, College of Graduate Studies

Date: 7/4/96

Dedicated to

my mother *Dr. (Mrs) Syeda Iqbal Akhtar*,
father *Dr. Hasan Arif*

&

my sister *Roumina Hasan*

whose prayers, guidance, and inspiration led to

this accomplishment

Acknowledgment

In the name of Allah, Most Gracious, Most Merciful. Read in the name of thy Lord and Cherisher, Who created. Created man from a {*leech-like* } clot. Read and thy Lord is Most Bountiful. He Who taught {the use of} the pen. Taught man that which he knew not. Nay, but man doth transgress all bounds. In that he looketh upon himself as self-sufficient. Verily, to thy Lord is the return {of all}.

(The Holy Quran, Surah 96)

First and foremost, all praise to Allah, *subhanahu-wa-ta'ala*, the Almighty, Who gave me an opportunity, courage and patience to carry out this work. I feel privileged to glorify His name in the sincerest way through this small accomplishment. I seek His mercy, favor, and forgiveness. And I ask Him to accept my little effort. May He, *subhanahu-wa-ta'ala*, guide us and the whole humanity to the right path *(Aameen)*.

Acknowledgement is due to King Fahd University of Petroleum & Minerals for providing support to this research work.

I am indebted to my thesis Chairman, Dr. Mayez Al-Mouhamed, for his kind help and advice. I acknowledge him for his valuable time, encouragement and guidance, especially during the early stages of this work. Working with him was indeed a pleasant and scholarly experience.

I am also profoundly grateful to my thesis Co-chairman, Dr. Habib Youssef, for his most sincere guidance and help in all matters. He displayed deep interest, put up constructive criticism and had stimulating discussions with me during the course of this work. In fact, I learned a lot more from him than just computer networks. Thanks are also due to my thesis committee member, Dr. Mostafa Abd-El Barr for his comments and critical review of the thesis.

I am thankful to the department chairman, Dr. Muhammad S. T. Benten, Dr. Samir Abdul Jauwad (ex-department chairman) and other faculty members for their co-operation.

I am thankful to my fellow graduate students and all my well-wishers and friends on the campus especially Kalim bhai, Pervez bhai, Abbi, Farook, Rehan, Nadeem, Ansar etc., who provided a wonderful company.

Last but not the least, thanks are due to the members of my family for their emotional and moral support throughout my academic career. No personal devel-

opment could ever take place without the proper guidance of parents. This work is dedicated to my parents for taking pains to fulfill my academic pursuits and shaping my personality decisively. They taught me the fundamental concept of life,

"*Tough times never last, tough people do*".

My mother deserves special mention for her consistent encouragement, inspiration, guidance and prayers. She has been instrumental in inculcating in me the self-confidence which is required to be successful in life. This acknowledgement would be incomplete if I don't mention my most beloved sister Roumina, and brother Shahab, who have always been at my side in the thick and thin and have made great sacrifices for me.

Contents

List of Tables

List of Figures

Abstract

Name: **Wasif Hasan**

Title: **A Novel Fast Packet Switch Architecture For ATM Networks**

Major Field: **Computer Engineering**

Date of Degree: **June, 1996**

In this thesis we present a novel fast packet switch architecture employing banyan interconnection networks, called the **Parallel-Tree Banyan Switching Fabric** *(PTBSF). It consists of a 3-D arrangement of banyan networks of the same topology in a tree structure. The packets enter the switch at the topmost level. Conflicting cells at the input of a switching element, in a particular stage, are routed vertically downwards to the corresponding switching element in the next level of banyan(s). Cell loss can occur only at the lowest level. The switch can be engineered to provide arbitrarily high throughput without the use of input buffering nor cell preprocessing prior to switching. The performance of the switch is evaluated analytically as well as through simulation under uniform traffic pattern and through simulation under a variety of ATM traffic workloads. The switch exhibited excellent performance with respect to cell loss and switching delay for all studied conditions compared to other similar architectures like the tandem banyan (TBSF)[5] and the piled banyan (PBSF)[6]. The advantages of PTBSF are high throughput, robustness under a variety of ATM traffic sources, low processing overhead and modularity.*

Master of Science Degree
King Fahd University of Petroleum and Minerals
Dhahran, Saudi Arabia
June, 1996

Chapter 1

Introduction

1.1 Overview

Currently most networks are dedicated to specific purposes like telephony, TV distribution, circuit-switched or packetized data transfer [1]. Using these pre-existing networks for the new emerging applications may lead to several drawbacks. The reason is that these networks are not adapted to the needs of the services which were unknown when they were implemented. Thus data transfer over the telephone network is hampered by the lack of bandwidth, flexibility and quality of analog voice transmission equipment.

In the 1960's, a worldwide effort began to upgrade public switched telephone networks (PSTN) from all-analog systems to systems supporting a combination of digital and analog signals. But in general the telephone networks were unsuited to

non-voice services to an extent that was required by the customer. Hence other dedicated networks came up like community antenna TV (CATV) network, telex networks, packet switched data networks (PSDN) etc. Each of these networks was specially designed for that specific service and most often totally unsuited to other applications.

The outcome of this service specialization is the existence today of a large number of private and public networks, quite independent of each other [7]. Each network requires its own separate design, manufacturing and maintenance in addition to the individual dimensioning according to the service type. Since resource pooling is impossible, each network must be designed to handle its worst case traffic condition. Moreover, these networks (especially the private ones) often employed non-standardized equipment, interfaces and protocols and are unable to offer access to other networks and users. Interface to the outside world is cumbersome and expensive. Thus, a need to build a network which can support different applications in an integrated fashion was felt. This laid the foundation for the introduction of narrow-band integrated services digital network (N-ISDN), in which voice and data are transported over a single medium. This was the first step towards the larger goal of a single universal network. But due to its limited bandwidth capabilities, N-ISDN cannot carry TV or video signals. Even the integration of narrow-band services like voice and data is rather limited [7].

Advances in VLSI and coding algorithms influence the bit rate generated by

a service and thus change the service requirement of the network. For example current digital N-ISDN switches are designed for 64 kbps voice channels. But with the current progress in speech coding and chip technology bit rates of 32 kbps or even lower will be used in future. Thus the internal available resources are used inefficiently. In future, new services with unknown requirements are expected to appear. A specialized network has great difficulties in adapting to changing or new service requirements rendering it inflexible.

Taking into consideration all these factors, namely, service dependence, flexibility and resource utilization, it is very important that in future we build a network which can handle all applications (existing and emerging) so that only a single service-independent network exists. This will enable sharing of all available resources between the different services. Moreover, there is a need to incorporate *broadband* features into ISDN because of the following reasons [1].

- The need to integrate *interactive* and *distributive* services and circuit and packet transfer modes into one universal broadband network.

- The emerging demand for broadband services like HDTV, video conferencing, high speed data transfer, video-telephony, video library, home education, video mail, high resolution graphics images and multimedia applications which will combine data, voice and video.

- The availability of high speed transmission, switching and signal processing

capabilities (bit rates of the order of hundreds of Mbps are being offered) due to advances in semiconductor and optical technology.

- Improved data and image processing capabilities are available to the users. Also there have been considerable advances in the software application processing in the computer and telecommunication industries.

The integration of services and incorporation of broadband features require the network to have the following qualities [7].

- The network should provide a common user-network interface for access to a variety of services including multiple bit rate traffic sources.
- The network should have *flexibility* and *scalability* such that it is able to adapt itself to changing or new needs.
- The network should be *efficient* in the use of its available resources by sharing them between all services.
- The network should be *cost effective*. Since only one network needs to be designed, manufactured, operated and maintained, the overall cost should be smaller.

Thus B-ISDN is conceived to become a universal network supporting different kinds of applications and customer categories. It supports switched semi-permanent and permanent, point to point and point to multipoint connections and provides on

demand reserved and permanent services. Connections in B-ISDN support both circuit switched and packet switched services of mono and multimedia type and of connection-less or connection-oriented nature in a unidirectional or bidirectional manner [1].

Asynchronous transfer mode (ATM) was chosen as the switching and multiplexing technique for B-ISDN. The ATM standard is designed to efficiently support high-speed digital voice and data communications. ATM is a circuit-oriented, hardware controlled, low-overhead concept of virtual channels (refer to section 1.2.2) which combines the advantageous features of both circuit switching and packet switching. Due to the small size of the packets and the high transfer rates offered by ATM, the transfer delay and the delay variations[1] are sufficiently low as in circuit switching. Moreover, due to the ability of ATM to multiplex and switch at the cell level, flexible bit rate allocation is possible as in packet switching.

In this thesis, we will mainly concentrate on the switching fabric used for ATM networks. Several architectures have been proposed for ATM switches. Those that have appeared in literature include shared memory architectures like the prelude switch [8], shared medium architectures like the ATOM switch [9], space division matrix architectures such as the knockout switch [10], space division architectures based on multistage interconnection networks such as the Batcher-Banyan network

[1]Delay variance is defined as the difference in time between two consecutive cells, arriving on the same switch input, to reach the switch output.

[11], [12], the Starlite switch [13], the Sunshine switch [14], the Tandem Banyan switching fabric [5], the Piled Banyan switching fabric [6] etc.

Multistage interconnection networks are among the most desirable building blocks for space division fast packet switches. There are several reasons for this, which will be discussed in section 4.2.3. The *Tandem Banyan switching fabric (TBSF)* is one such architecture with MINs as the basic building blocks. It consists of placing Banyan networks in series in order to provide multiple paths from each input to each output thus avoiding the internal blocking to which Banyan networks are susceptible and also achieving output buffering. But the delay variance in this network is large and the packets may get out-of-sequence when the size of the network increases. It is also not very adept at handling bursty traffic. Thus, our endeavor in this work was to overcome these limitations by investigating new architectures.

1.2 Asynchronous Transfer Mode

ATM is considered to be the ground on which B-ISDN is to be built [1]. The term *transfer mode* is a specific way of transmitting and switching information in a network. In 1988, ATM was selected as the switching and multiplexing technique for B-ISDN. The impetus for ATM is based on the following reason [2].

> *Networks with high bandwidth and low latency are required for real time transmission capabilities necessary for future multimedia applications. For an ap-*

plication requiring the transmission of large amounts of data in real time, new network architectures and protocols must be designed which support multiple service classes of data in an efficient and cost effective manner.

In ATM based networks, the multiplexing and switching of packets is independent of the actual application. So the same equipment can handle a low as well as a high bit rate connection, be it of stream or burst nature. Dynamic bandwidth allocation on demand with a fine degree of granularity is provided. Whereas today's networks are characterized by the coexistence of both circuit switching and packet switching, B-ISDN will rely on a single new method, ATM, which combines the advantageous features of both circuit and packet switching techniques [1]. The former requires low overhead and processing and once a connection is established, the transfer delay of the information being carried over it is low. The latter is much more flexible in terms of the bit rate assigned to virtual connections. ATM is a circuit-oriented, hardware controlled, low-overhead concept of virtual channels which have no flow control or error recovery. The ATM model has several objectives [15], as listed below.

1. ATM must be scalable and cost effective.

2. It must support applications with diverse traffic requirements (example, multimedia applications) and be able to support multiple data streams with acceptable delay bounds.

3. ATM must be able to perform multicast operation efficiently, as many emerging applications will require frequent use of this kind of operation.

4. ATM must be inter-operable with existing LANs, MANs and WANs and should use existing standards and protocols whenever possible.

Many new applications are expected to take advantage of the flexibility in switching and high speed that ATM provides. For example, digital medical imaging may be one of the first applications to utilize a network based on ATM [16]. A second application area that will benefit from ATM networks is that of providing supercomputer support to research centers. A third potential application is the interconnection of powerful workstations which are geographically distributed. These computers can be used together to cooperatively solve problems previously requiring large and costly supercomputers.

High bandwidth applications such as these, will generate a wide diversity of traffic which is very difficult to support efficiently with existing networks. Thus, there is need for ATM technology which can support data traffic with widely varying service requirements. Contrary to other networking techniques, ATM needs minimal network node functionality and thus allows very high network speeds and low delay to be attained.

Figure 1.1: ATM cell structure.

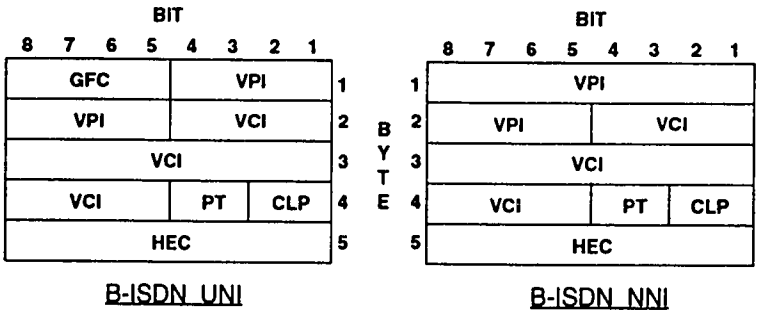

Figure 1.2: Cell header at B-ISDN UNI and B-ISDN NNI.

1.2.1 ATM Cell Structure

ATM is based on a virtual connection-oriented packet (or cell) switching mode. In ATM all information to be transmitted is broken into fixed size data blocks called *cells*. A cell consists of five bytes of header and 48-bytes of information field as shown in Fig. 1.1. The header field contains control information such as identification, cell loss priority, routing and switching information for the cell. The cell header is shown in detail in Fig. 1.2.

The cell header at the B-ISDN user-network interface (UNI) is different from that at the network-node interface (NNI) in the use of the bits 5-8 of byte 1. The NNI is the interface between network nodes. At NNI these bits are part of the VPI,

while at UNI these bits form an independent unit called *generic flow control (GFC).* The various fields within the header and their functions are described below.

Generic Flow Control (GFC):

The GFC field consists of 4 bits. It is used by the UNI to control the amount of traffic entering the network. This allows the UNI to limit the amount of data entering the network at times of congestion. The GFC mechanism supports both point to point and point to multipoint configurations.

The ATM network does not provide the type of flow control which is implemented in packet networks and it does not have the facility to store cells for a long period of time. So inside an ATM network there is no need for GFC. GFC only controls terminals connected to a customer network.

Virtual Path Identifier/Virtual Channel Identifier (VPI/VCI):

The *virtual path identifier (VPI)* field at the B-ISDN UNI consists of 8 bits and is used for routing. The VPI at the NNI comprises the first 12 bits of the cell header, thus providing enhanced routing capabilities. For some special purposes, pre-assigned VPI values are used. The *virtual channel identifier (VCI)* field consists of 16 bits in both UNI and NNI and is used for channel identification. It also has some pre-assigned values.

Payload Type Indicator (PTI):

Two bits are used for this field. The PTI field is used to distinguish between user cells and control cells. This permits the control and signalling data to be transmit-

ted on a different subchannel from user data.

Cell Loss Priority (CLP):

The CLP field indicate explicitly the priority of the cell being lost i.e., this field is used to indicate whether a cell may be discarded during periods of network congestion. For example, voice data may be able to suffer lost cells without the need for retransmission, whereas text data cannot. Thus an application may assign a higher CLP for voice traffic.

Header Error Control (HEC):

This field is used to protect the header from transmission errors.

1.2.2 ATM Cell Switching

The VPI/VCI fields together are used to determine which cells belong to a particular connection. The VPI and VCI values are used by the routing protocol to determine the path(s) and channel(s) a cell will travel through. Fig. 1.3 shows the relationship between virtual channel, virtual path and transmission path. A transmission path may consist of several virtual paths and each virtual path may carry several virtual channels. A concatenation of VC links is called *virtual channel connection (VCC)*, and a concatenation of VP links is called *virtual path connection (VPC)* [1].

On each incoming link of an ATM switch, an arriving cell's VPI and VCI values uniquely determine the new virtual identifier to be placed in the cell header and

Figure 1.3: Relationship between virtual channel, virtual path and transmission path.

the outgoing link over which to transmit the cell. Two hosts can use a virtual path to multiplex many individual application streams together, using the VCI to distinguish between these streams. The virtual path concept was incorporated to provide the capability to manipulate a set of ATM connections in one unique channel [17].

The manner in which ATM carries out cell switching is shown in Fig. 1.4 [2]. At the time of connection set up between two or more hosts on the network, a virtual path is defined between the source and the destination. Internal routing tables in the switches are initialized by the connection establishment procedure. Upon entering an ATM switch, a cell's VPI field is used as an index to the routing table that determines which output port (destination) the cell should be routed to. At the same time a *new* VPI value may be placed in the cell and the cell forwarded to the next switch. The VPI's are used to establish virtual paths on a semipermanent basis between network endpoints while the VCI's are used to establish virtual links over a given virtual path connection. The network does not modify the VCI fields of cells on virtual path connections. Thus, the hosts can set up new virtual channels

on an established virtual path, on their own, without requesting the network.

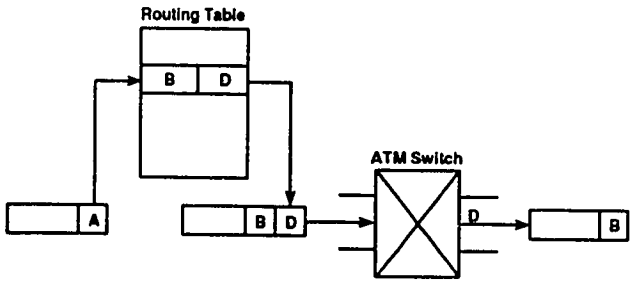

Figure 1.4: ATM cell switching[2].

The cells belonging to a connection do not always have to be transmitted in a continuous manner, one after the other, in the data stream. The cells are, in fact, statistically multiplexed with the amount of bandwidth allocated to a connection determined by the traffic requirement of the connection. ATM allows very efficient utilization of network bandwidth, as statistical multiplexing allows the total available bandwidth to be dynamically distributed among a variety of user applications. This is done by providing the label field inside the ATM cell header (VPI and VCI) and thus selecting the virtual channel paths according to the anticipated traffic and allocating the network resources needed. Thus ATM can provide communication with a bit rate that is individually tailored to the actual need, including time varying bit rates, just like packet switching techniques. Fig 1.5 shows how the channels are allocated.

In *synchronous transfer mode (STM)*, a packet is identified by its position in

the transmission frame, while in ATM a cell or a packet associated with a specific virtual channel may occur at virtually any position as shown in Fig. 1.5.

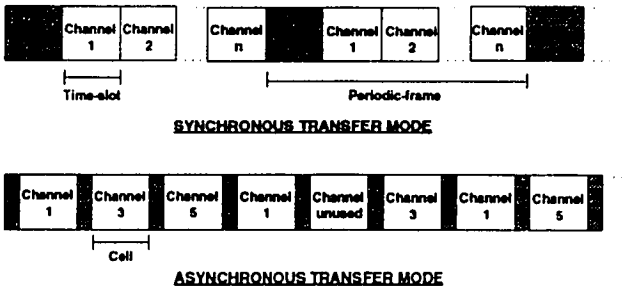

Figure 1.5: Difference between STM and ATM[1].

1.2.3 ATM Protocol Reference Model

The functions of the layers and the relation between the layers are described in the *protocol reference model (PRM)* [1]. The B-ISDN PRM consists of the following three layers as shown in Fig. 1.6.

- User Plane
- Control Plane
- Management Plane

The *management plane* includes two types of functions called *layer management* and *plane management* functions. All the management functions pertaining to the whole system are located in the plane management. Its task is to provide coordination

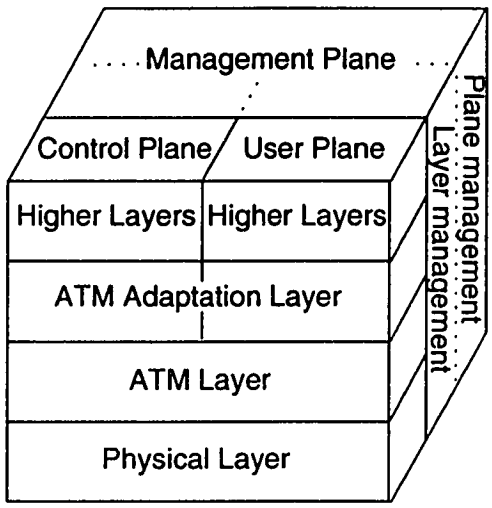

Figure 1.6: B-ISDN Protocol reference model[1].

between all planes. Within this plane there is no layered structure. The layer management has a layered structure. It performs the management functions relating to resources and parameters residing in its protocol entities. For each layer, the layer management handles the specific operation and maintenance information flows.

The *user plane* handles the transfer of user information. All associated mechanisms like flow control and recovery from errors are included. User plane is layered.

The control plane is responsible for call control and connection control functions. These are the signalling functions which are necessary to set-up, supervise and release a call or a connection.

The functions of the *physical layer (PL)* and the ATM layer are the same for the control plane and the user plane. But the functions of the ATM adaptation

layer and the higher layers may be different for the two planes. The functions of the various layers are shown in Table 1.1.

	HIGHER LAYER FUNCTIONS	HIGHER LAYERS	
LAYER MANAGEMENT	CONVERGENCE	CS	AAL
	SEGMENTATION AND REASSEMBLY	SAR	
	GENERIC FLOW CONTROL CELL HEADER GENERATION/EXTRACTION CELL VPI / VCI TRANSLATION CELL MULTIPLEX DEMULTIPLEX	ATM	
	CELL RATE DECOUPLING HEC SEQUENCE GENERATION / VERIFICATION CELL DELINEATION TRANSMISSION FRAME ADAPTATION TRANSMISSION FRAME GENERATION / RECOVERY	TC	PHYSICAL LAYER
	BIT TIMING PHYSICAL MEDIUM	PM	

Table 1.1: The functions of B-ISDN in relation to B-ISDN PRM [1].

The Physical Layer:

The *physical layer (PL)* encodes and decodes the data into suitable optical/electrical signals for transmission and reception on the communication medium used. The physical layer also provides cell delineation functions, header error control (HEC) generation and processing, performance monitoring and payload rate matching of the different transport formats used at this layer.

The transmission of ATM cells from one user to another can be done by the PL in two ways. At the *user network interface* (UNI), ATM cells may be carried in an

externally framed synchronous transmission structure or in a cell based asynchronous transmission structure. The optical data rates, synchronization and framing format chosen for B-ISDN are called the *synchronous digital hierarchy (SDH)* in Europe and the *synchronous optical network (SONET)* in North America. The basic time unit of a SONET frame is 125 microseconds. Fig. 1.7 shows the SONET frame structure [2].

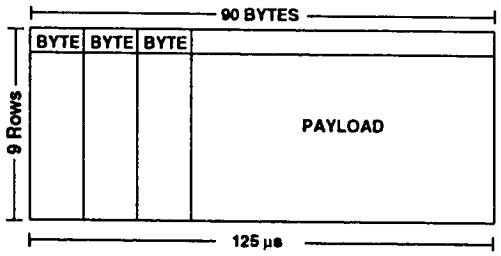

Figure 1.7: SONET STS-1 payload envelope.

The basic building block of SONET is *synchronous transport signal level 1* (STS-1) with a bit rate of 51.84 Mbps. The STS-1 frame structure consists of 90 columns and 9 rows of 8-bit bytes (Fig. 1.7). The order of transmission of bytes is row by row from left to right, with one entire frame transmitted every $125\mu s$. The first three sections of STS-1 contain section and line overhead bytes used for error monitoring, system maintenance functions, synchronization and identification of payload type. The remaining are used to carry the payload. Higher-rate SONET signals are obtained by byte-interleaving n frame-aligned STS-1's to form an STS-n. For example, STS-3 has a bit rate of 155.52 Mbps and STS-12 has a bit rate of 622.08 Mbps.

The ATM Layer:

The ATM layer is a unique layer that carries all the different classes of services supported by B-ISDN within a 53-byte cell. In the transmit direction this layer is responsible for multiplexing cells from individual VPs and VCs into one cell stream by the *cell multiplexing* function. At the receiving side the *cell demultiplexing* function splits the arriving cell stream into the appropriate VPs and VCs. This layer also performs cell rate decoupling or unassigned cell insertion and removal, priority processing and scheduling of cells, cell loss priority marking and reduction, cell rate pacing and peak rate enforcement, explicit forward congestion marking and indication, cell payload type marking and differentiation and generic flow control access. The functionality of the ATM layer is defined by the fields present in the ATM cell header (see Fig. 1.2) except for the HEC generation and verification which is the function of the PL.

In order for several ATM channels to be supported in a single SONET STS-n frame, the rate of valid cells must be adapted to the capacity of the transmission payload. To achieve proper alignment, idle bytes are inserted and extracted into the synchronous frame structure at the endpoints of the network. A pointer carried in the overhead bytes of the STS header is used to indicate the position of the cell within the payload frame. Thus, cells do not have to be strictly frame-aligned with the underlying payload signal. That is why we refer to this mode of transmission

as *asynchronous transfer mode.* In order to identify a certain time slot, we need to have an additional header field containing a VPI and VCI.

The ATM Adaptation Layer (AAL):

The basic function of this layer is to provide a link between the services provided by the ATM layer and the requirements of the higher layer. Higher layer *protocol data units* (PDUs) (the unit of data in an N layer peer-peer protocol is called protocol data unit) are mapped into the information field of an ATM cell.

The AAL layer consists of two sublayers; *segmentation and reassembly sublayer (SAR)* and the *convergence sublayer (CS).* The SAR sublayer performs the segmentation of higher layer PDUs into a suitable size for the information field of the ATM cell (48 bytes) at the transmitting side and reassembly of the particular information fields into the higher layer PDUs at the receiving side. The CS is service dependent and provides the AAL service at the AAL-SAP (service access point). The point at which the layer N services can be accessed by the layer above is called *N-service access point.* The AAL also plays a key role in the internetworking of different networks and services.

To minimize the number of AAL protocols four service classes have been defined. The classification has been performed according to the following parameters.

1. timing relation between the source and the destination

2. bit rate

3. connection mode

Fig. 1.8 shows the various AAL service classes[1]. The service classes are :

1. **Class A:** A time relation exists between the source and the destination, the bit rate is constant and the service is connection-oriented. For example, voice and constant bit rate video.

2. **Class B:** A time relation exists between the source and the destination, the bit rate is variable and the service is connection-oriented. For example, variable bit rate audio and video.

3. **Class C:** No time relation exists between the source and the destination, the bit rate is variable and the service is connection-oriented. For example, a connection-oriented data transfer.

4. **Class D:** No time relation exists between the source and the destination, the bit rate is variable and the service is connection-less. For example, LANs interconnection, e-mail etc.

Initially four types of AAL protocols were proposed to support the four service classes defined, namely, type 1, 2, 3 and 4. Type 3 and 4 were later merged into a single type (called AAL type 3/4), since the differences between them were minor. A fifth AAL type has also been proposed due to the high complexity of the AAL

	CLASS A	CLASS B	CLASS C	CLASS D
TIMING RELATION BETWEEN SOURCE AND DESTINATION	REQUIRED		NOT REQUIRED	
BIT-RATE	CONSTANT	VARIABLE		
CONNECTION MODE	CONNECTION ORIENTED			CONNEC-TIONLESS

Figure 1.8: Service classification for AAL[2].

type 3/4 [18], [19]. The AAL type 5 protocol is sometimes called the simple and efficient adaptation layer (SEAL).

1.3 ATM Switch Architecture

An ATM switch consists of a set of N input and N output ports, a switch fabric and a management and control processor (MCP) as shown in Fig. 1.9 [3].

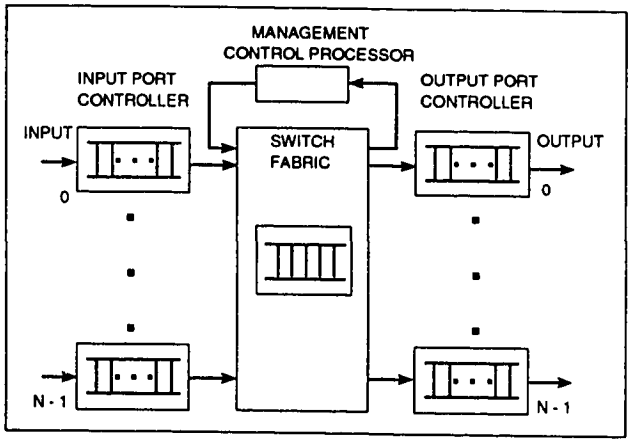

Figure 1.9: Model of an ATM switch[3].

Input-Output Controllers:

Each port is managed by an intelligent controller. Input port controllers are responsible for providing buffering, cell duplication, multicasting, cell processing, VCI translation, multiplexing traffic from several low-speed devices and path connection requests and reservations through the switch fabric. Likewise, the output controllers can provide buffering, VCI translation, demultiplexing and an N:R selection (selecting R packets out of a maximum of N for buffering), where R is the queue limit for each output port.

Switch Fabric:

The switch fabric is essentially a device which routes cells from input to output ports. Its basic functions are:

- establishing a path between an input and output port within the switch
- service discipline for input ports
- resolution of internal blocking
- supporting several input-output connections simultaneously

Management and Control Processor:

MCP is responsible for communicating with the port controllers and supervising the switch operation, administration and management.

1.4 Conclusion

ATM is emerging to be the technology of choice for implementing future B-ISDN. It combines the advantageous features of both circuit-switching and packet-switching. In this chapter, we discussed the factors which lead to the development of B-ISDN and the reasons for the acceptance of ATM as the switching and multiplexing technique for B-ISDN. We also presented an overview of the ATM technology, its cell structure, its protocol reference model and the method by which cells are switched in ATM.

ATM being a switch based technology, central to its success is the switch architecture. Any switch proposed for ATM must support traffic sources with diverse traffic characteristics. It must also have high throughput, low latency and must be cost effective. In the next chapter, we will discuss the various switch architectures which have been proposed for ATM networks.

Chapter 2

Literature Survey

2.1 Introduction

This chapter is a detailed survey of the various switch architectures which have been reported in literature for ATM networks. The chapter begins with a classification of the switch architectures, based on the physical connection between the input and output ports of the switch fabric. The switches have been divided into two broad classes: time division and space division architectures. Shared memory and shared medium switches fall into the time division architecture category. Switches such as crossbar, bus matrix and those based on multistage interconnection networks (MINs) are examples of space division architecture. Because of the importance of switches based on MINs, we discuss them in a separate section.

2.2 Classification of Switch Architectures

The switch architectures can be classified based on different features of the architecture. The most commonly used classification is based on the physical connection between the input and output ports within the switch fabric [3]. Based on this criterion, the switch architectures can be of time or space switching type. The classification is shown in Fig. 2.1.

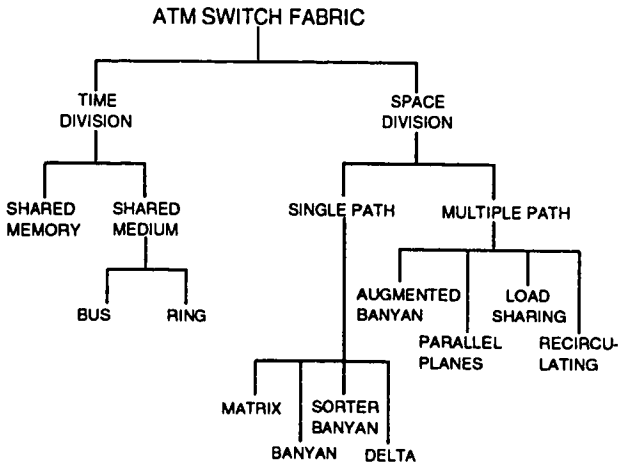

Figure 2.1: A taxonomy of the ATM switch fabric[3].

In *time division switches*, the physical resource (say, a conducting medium or a memory) is time shared among several input-output connections. For example, bus is a physical conducting medium in which time division multiplexing can be applied. A memory module can also implement the time division multiplexed connections. A memory module can be used to store cells coming from the input ports which

can then be accessed by the output ports. In order to implement a shared memory architecture efficiently, a multiported memory with up to $2N$ ports is required. The cost and complexity of the memory is proportional to N.

In *space division switches*, the switch fabric can support multiple connections simultaneously. The connections are based on the availability of non-conflicting paths within the fabric. For example, $2X2$ switches can be arranged in a grid structure to form a space division switch. Switches based on space division can provide single or multiple paths between a pair of input and output ports. There are two ways to route a cell through a space-division switch: 1) self routing and, 2)label routing [20]. In *self routing*, the switching fabric itself takes care of routing a cell as the routing relies on the regular interconnection of the switching elements within the fabric. In *label routing*, the VCI field within the header is used by each switching element to decide the output link onto which the cell has to be forwarded.

Time and space division switching can be combined such that several time-division switches are interconnected via a space-division switch in a hierarchical fashion [3]. Time division strategy can be deployed at the interface to the ATM switch and space division inside the switch. At the interface, the time division architecture will support expandability and non-uniformity of port data rates while inside the switch the space division architecture will support parallelism and speed.

2.3 Time Division Fast Packet Switches

As mentioned earlier, *time division switches* employ time division multiplexing in order to share a physical resource (bus or a memory) among the input and output ports. Time division switches can be either of the *shared memory* or the *shared medium* type.

2.3.1 Shared Memory Switches

The shared memory switches are the most natural type of fast packet switches because of their similarity to traditional packet switches in WANs. The switch consists of a single dual ported memory shared by all input and output lines. Packets arriving on all input lines are multiplexed into a single stream which is fed to the common memory for storage. Internally to the memory, packets are arranged into separate output queues, one for each output line. Simultaneously an output stream of packets is formed by retrieving packets from output queues sequentially, one per queue. The output stream is then multiplexed and packets are transmitted on the output lines [4]. Fig. 2.2 shows a shared memory switching architecture.

Efficient implementation of the shared memory approach requires a multiported memory with up to $2N$ ports. The cost and complexity of the memory is proportional to N. An example of the shared memory switch is the Prelude switch [8], developed at CNET.

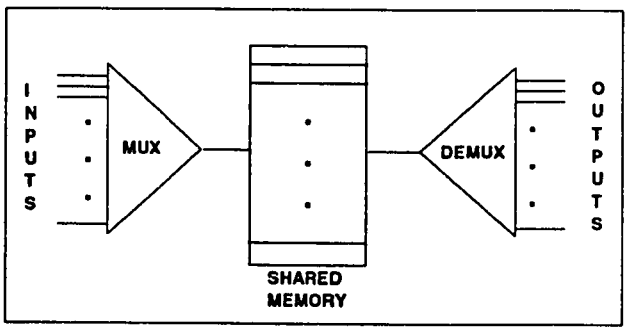

Figure 2.2: Shared memory switching architecture.

2.3.2 Shared Medium Switches

A *shared medium* switch also involves multiplexing of the incoming packets into a single stream and then demultiplexing them into individual streams, one for each output line. In shared medium switches, all packets arriving on the input lines are synchronously multiplexed onto a common high speed medium, generally a bus. The bus has a bandwidth equal to N times the single input lines, where N is the number of inputs of the switch. Each output line is connected to the bus via an interface consisting of an address filter (AF) and an output FIFO buffer. Depending on the packet's output address, the address filter of a particular output line determines whether that packet is to be stored into the FIFO buffer or not [4]. The basic structure of the shared bus architecture is shown in Fig. 2.3.

The single path through which all packets flow here is broadcast time-division bus and the demultiplexing is basically done by the address filters in the output

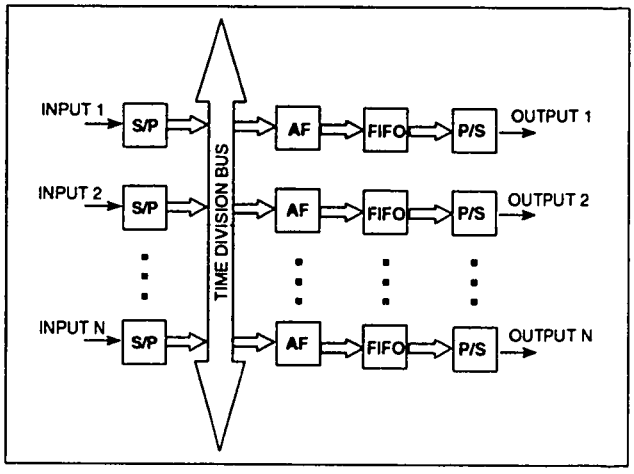

Figure 2.3: Basic structure of a shared-bus type architecture[4].

interfaces. An example of the shared medium switch is the Atom switch [9].

2.4 Space Division Fast Packet Switches

In a space division switch, multiple concurrent paths are established from the inputs to the outputs, each with the same data rate as the individual line. Hence no component in the switching fabric has to run at a speed higher than $2V$, where V is the line speed. Another distinct feature is that the control of the switch need not be centralized, but may be distributed throughout the fabric [4].

2.4.1 Crossbar Switching Fabric

A *crossbar switch* [21] with N inputs, consists of a square array of N^2 cross-point switches, one for each input-output pair. Closing the (i, j) cross-point switch, establishes a physical connection between input line i and output line j. The N^2 cross-point switches in the crossbar fabric enable N pairs of input/output connections to be established simultaneously, provided that no two destinations are the same. Thus, there is no internal blocking in crossbar switches. But their complexity increases with $O(N^2)$, placing a limit on their size and scalability. Fig. 2.4 shows a crossbar switching fabric.

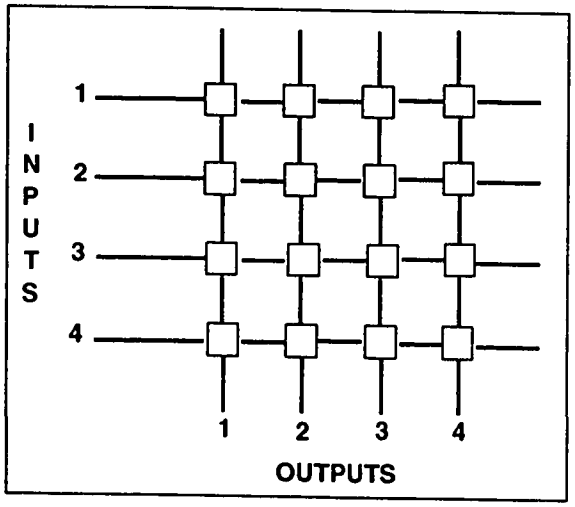

Figure 2.4: Crossbar switching fabric.

2.4.2 Bus Matrix Switching Fabric

The *Knockout switch* [10], is an example of a bus matrix architecture. It consists of N broadcasting buses. The bus interfaced at input port i can transmit to all the output ports. The switch has a concentrator at each output port and N filters. The N packet filters at each output port filter the packets off the buses that are destined to that port. The switch is self-routing because the filtering is based on the destination address. Fig. 2.5 shows the knockout switch.

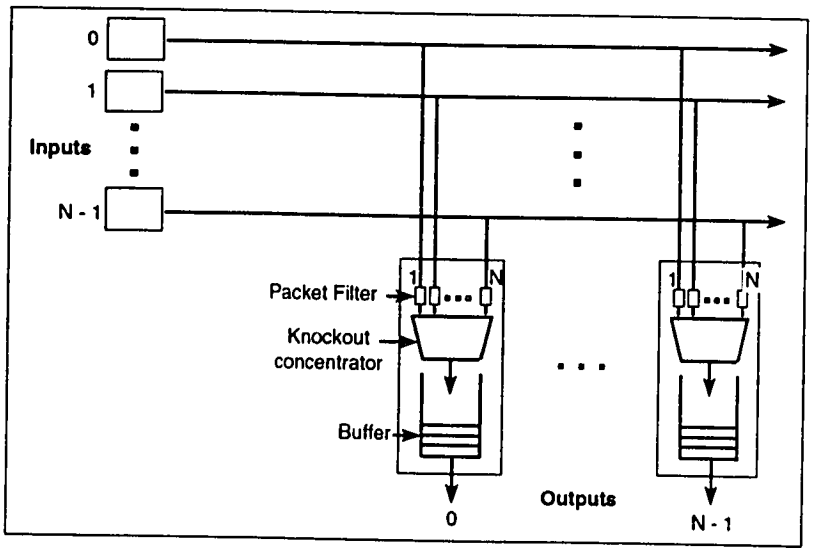

Figure 2.5: The Knockout switch.

Each output port has B buffers. The function of the concentrator is to select B packets out of the maximum of N packets arriving at that port. Thus $N-B$ packets may get lost in the worst case. The concentrator has B sections of competition with B rounds in each section. The end of each section results in the assignment of one

of the output buffers. During the first round of the first section, N packets contend for the first output buffer. After the first round, $N/2$ packets are left behind. The winners move to the next round at the end of which $N/4$ packets will be left behind. This process continues till two packets contend for the first output buffer in the last round. Then $N-1$ losers of the first section contend for the second buffer in the next section and so on till all the B buffers are assigned.

2.5 Switches Employing MINs

Although switches employing MINS are also space division switches, but due to their special significance as ATM switches we will discuss them separately in this section. MINs are among the most desirable building blocks for space division fast packet switches. The reasons for this are listed below [5].

1. MINs are space division switches with multiple concurrent paths from the inputs to the outputs.

2. The control of the switch need not be centralized, but may be distributed throughout the switching fabric.

3. The complexity of a crossbar switch (also a space division switch) increases with $O(N^2)$, placing a limit on its size and scalability. The complexity of a MIN, for example, a Banyan network is only O(Nlog(N)).

4. Paths can be established from the input to the output using a self routing procedure.

5. MINs possess a regular structure which is particularly suitable for VLSI implementation.

6. Lastly, the structure of MINs is modular, allowing the building of large networks from smaller ones without the need to modify considerably the physical layout and the algorithm used in their operation, i.e., they are easily *scalable*. Scalability is one of the most desirable features of an ATM switch.

They do have some drawbacks. These are the following.

- Depending on the particular internal fabric used and the resources available within, it may not be possible to establish all the required paths simultaneously due to internal conflict on a link. This is referred to as *blocking* within the network and induces throughput limitations.

- Since there is no coordination among the arriving packets, it may be possible for more than one packet arriving in the same time slot, to be destined to the same output port. This is referred to as *output conflict*.

Thus all switch architectures based on Banyan networks employ means to overcome blocking and output conflict in order to improve throughput.

A MIN can be constructed from simple switching elements arranged in multiple stages. Fig. 2.6 shows a switching element and its states. Packet switch architectures

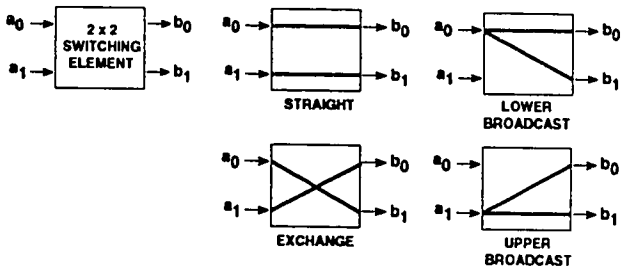

Figure 2.6: Switching elements and their states.

used MINs such as Banyan and Delta. These are $N \times N$ switches composed of $b \times b$ simpler switching elements. Generally they consist of K stages, where $N = b^K$ (or $K = log_b N$) and N/b switches in each stage. As mentioned earlier these networks are susceptible to internal blocking and output conflicts. There are several ways to overcome these limitations and increase the throughput [22].

- Adding a sorting network in front of the switch to reduce or eliminate the possibility of internal blocking.
- Recirculate packets that cause blocking, back to the input ports and include them in the next cycle.
- Provide a contention-resolution phase among the input ports.
- Add extra stages to the multistage network in order to provide multiple paths between each source and destination.
- Provide buffering on each link or switching element.

- Increase the network throughput by using several networks in parallel.

We will present here some well known paradigm architectures which employ a combination of some of the above mentioned techniques in order to achieve high throughput.

2.5.1 Switches Employing Recirculation

The Starlite Switch:

The *Starlite switch* [13], employs recirculation to avoid output blocking in Batcher-Banyan networks [11]. A trap network is placed between the Sorter and the routing Banyan network called the expander (see Fig. 2.7). The trap network removes all but one common destination packets from the output of the Sorter network and routes them back to the Sorter network for transmission in the next cycle. The routed-back packets are placed on the empty input ports. An aging (time-stamp) mechanism gives the old packets higher priority for inclusion in the next sort group.

The main drawback of this architecture is that recirculation results in sequencing problem. In order to overcome this problem, Hui and Arthurs [23] proposed a three phase algorithm which includes a contention resolution phase for external conflicts. But the feedback links are still there, which restrict the size of the network.

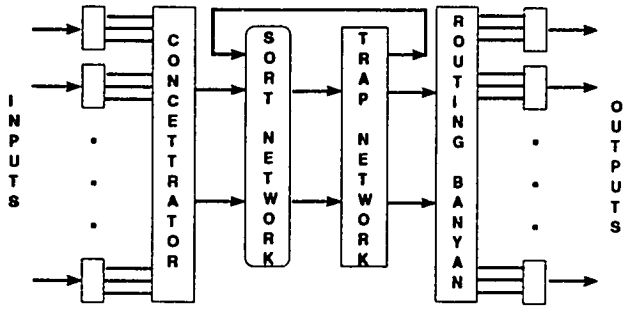

Figure 2.7: Starlite switch architecture.

The Sunshine Switch:

The *Sunshine switch* [14], employs several of the above mentioned techniques in order to increase the throughput. The architecture (shown in Fig. 2.8) consists of a Batcher sorting network at the input. The output of the Batcher network consists of cells sorted on the basis of their destination address. The trap network which follows, selects at most K cells per destination address to be routed. The remaining cells are separated by the concentrator and forwarded to the recirculation buffer. The recirculation buffer delivers these cells to dedicated input ports for transmission in the next cycle. The recirculation buffer consists of T parallel paths to the input of the Batcher network with one unit of delay. At the output there are K parallel Banyan networks. The selector delivers the cells to the Banyans to be routed. Each input and output port is managed by a controller, IPC and OPC respectively.

Here again, recirculation results in sequencing problem and the feedback links restrict the size of the network. Moreover, there is considerable amount of prepro-

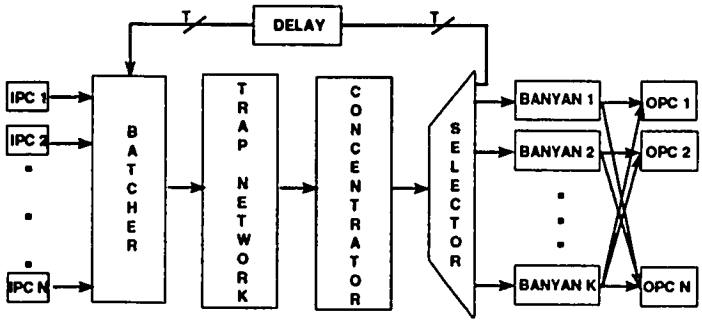

Figure 2.8: Sunshine switch architecture.

cessing before the cells are finally submitted to the switching fabric.

2.5.2 Switches Employing Multiple Outlets

Multistage interconnection networks with multiple outlets can pass a number of packets having the same destination to the output as opposed to non-blocking networks in which although there are no internal conflicts, output conflicts still restrict the throughput to around 63% [6]. Thus, MINs with multiple outputs can support much higher throughput, as is required in ATM networks.

Multi Banyan Switch:

The simplest of the MINs with multiple outputs is the *multi banyan switching fabric* (MBSF) [24]. This architecture has been evaluated as a circuit switching and ATM packet switching network [25], [26]. It consists of a number of homogeneous Banyan planes arranged in a parallel 3-D arrangement as shown in Fig 2.9. Each input

port can distribute its traffic to a number of Banyan planes of the same type (same interconnection function). The output from any plane can feed a packet to the corresponding output port. The throughput increases very slowly with the number of Banyans and never reaches *one*.

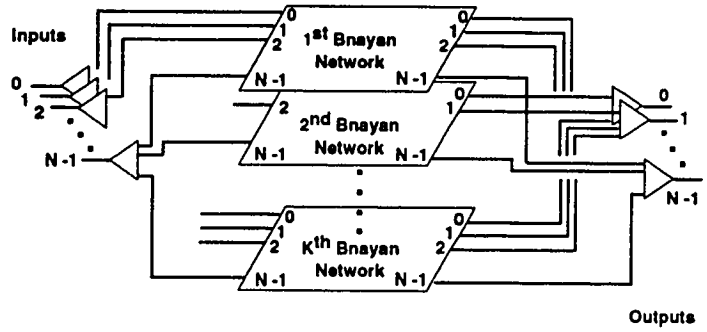

Figure 2.9: Multi banyan switch architecture (MBSF).

Expanded Banyan Switch:

The *expanded banyan switching fabric* (EBSF) [27] is based on an expanded delta network. An ($N \times N$) expanded delta network with an expansion factor, EF, can be constructed by interleaving EF number of ($N \times N$) delta networks as shown in Fig 2.10 for $N = 4$ and $EF = 3$. The resulting network is a ($P \times P$) interconnection network, where $P = N \times EF$, and consisting of $log_2 N$ stages. Each stage has $P/2$ switching elements. In a given time slot, only N links are utilized. Each output port can accept packets from EF switching elements in the last stage. Thus, increasing EF leads to both decrease of internal blocking as well

as output contention (as more cells can be simultaneously transferred to a given destination).

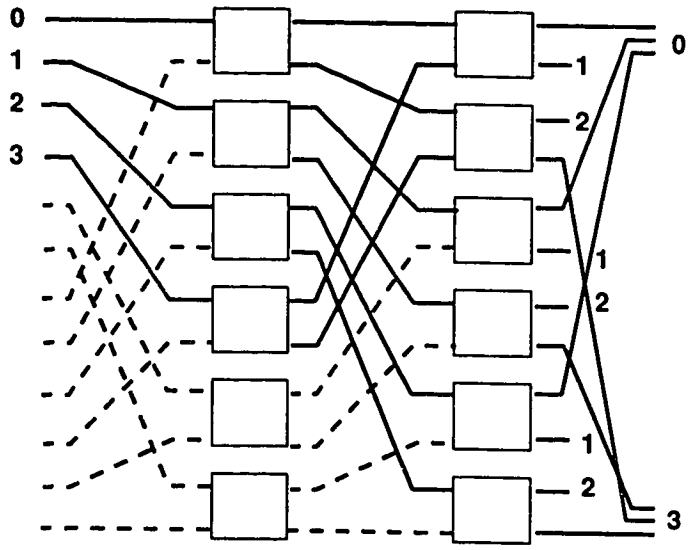

Figure 2.10: Expanded banyan switch architecture (EBSF).

In [27], it has been shown that for a switch size of 1024 and $EF = 16$, the expanded banyan switch can achieve a throughput of above 90%. Hanawa et. al.[6] compared the performance of EBSF with that of other banyan based architectures like MBSF, their own proposed architecture, the piled banyan switch and the tandem banyan switch proposed by Tobagi et. al.[5]. They showed that although EBSF performs better than MBSF, its performance as compared to tandem and piled is far below for a switch size of $N = 256$ with equivalent amount of hardware.

Tandem Banyan Switch:

The *tandem Banyan switching fabric* (TBSF) [5] consists of a number of Banyan networks (say K) in series, such that each output of every Banyan network is connected to both the corresponding input of the next network in series and the corresponding output buffer except the last Banyan for which the outputs are only connected to the output buffers. Fig. 2.11 shows the tandem Banyan switching fabric.

When a conflict between two packets occurs at some switching element, one of the two packets (depending on the priority) is routed to the correct output link while the other is misrouted. Also whenever a packet gets misrouted it is not allowed to interfere with the routing of correctly routed packets within that network, in future. At the output of the first Banyan, those packets that have reached their correct destinations are delivered to the corresponding output buffers, while the packets which have been misrouted are fed into the second Banyan for further processing. This process is repeated for K networks in series. Unsuccessful packets at the output of the last stage get lost. Since the load on successive Banyan networks decreases, hence the probability of conflicts also decreases.

Tobagi et. al. analyzed the performance of this architecture under a variety of traffic conditions. TBSF can achieve a throughput close to one if the number of banyans in series is large under most traffic conditions. But it has the following drawbacks:

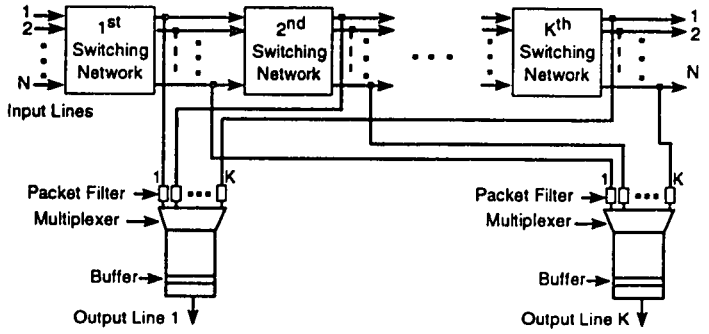

Figure 2.11: Tandem banyan switching fabric (TBSF).

- Possibility of the packets being delivered out-of-sequence if the condition $(K-1)\tau_b > T$ is satisfied, where K is the number of banyans in series, τ_b is the delay within a banyan, and T is the slot size in seconds [5]. To deal with the out-of-sequence packet delivery problem all packets reaching the buffer from Banyan network i are delayed by $(K-i)\tau_b$. Thus all packets in a given time slot arrive at the buffer in time $t + K\tau_b$. This will quite obviously lead to slowing down of the whole network considerably and places a *limit* on the *size* and *scalability* of the network.

- Due to the random delay between the packets on a given input line in being routed to the output, the *variance of delay* between packets is likely to be high.

- In TBSF when we misroute a packet at some stage, say i, that packet has to be routed from the first stage onwards in the next banyan. Thus, the time

spent in routing the packet correctly upto i^{th} stage is wasted. This leads to higher average delay.

Piled Banyan Switch:

Hanawa et. al. [6] proposed the *piled banyan switching fabric* (PBSF). PBSF consists of arranging banyan networks one above the other at different levels as shown in Fig 2.12. The packets can be routed from one banyan plane to the next one below it through the vertical links between the banyan planes. The packets arrive at the input of the banyan network in the first level. On conflict, one of the packets is routed vertically downward to the corresponding switching element of the next plane. From the 2^{nd} level onwards, the packet at the vertical input may conflict with the packets at the horizontal inputs. Thus, if three packets collide, the packet at the vertical input is routed to the next stage in the same banyan, one horizontal packet is routed to the next lower layer banyan while the third packet is lost.

The main drawback of this architecture is that the loss can occur anywhere except in the top level. The *throughput* of PBSF remains saturated at 98%, even if a large number of banyans is used. This is due to the fact that cell loss which has already occurred in the upper layers cannot be compensated for in the lower layers.

In this thesis, we present a novel banyan based switch architecture called the *Parallel-Tree Banyan Switching Fabric* (PTBSF), which has the following desirable characteristics:

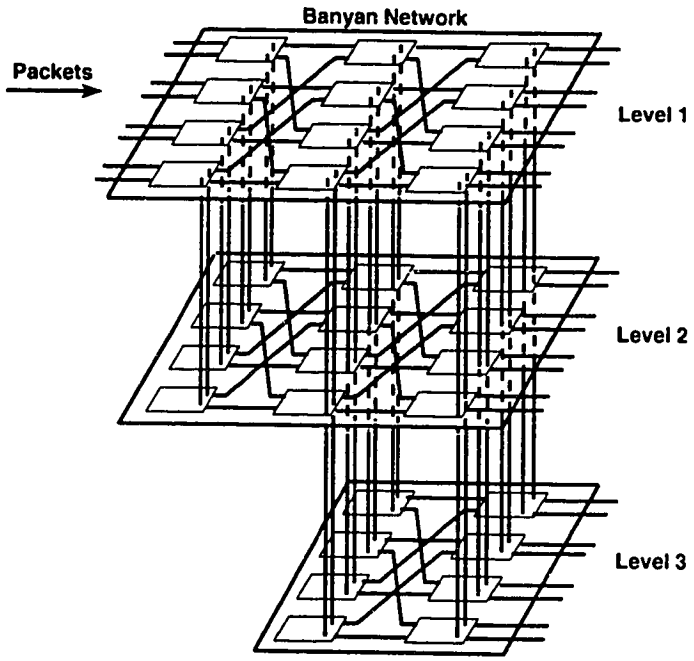

Figure 2.12: Piled banyan switching fabric (PBSF).

1. The switch performs straightforward routing with no buffering nor recirculation to minimize processing time (switching delay) and hardware complexity.

2. The performance of the switch scales up very well with the addition of extra hardware resources. Moreover, cell loss rate can be reduced to an arbitrarily low value (even zero) if we are willing to pay the cost.

3. The switching delay is very small compared to other reported architectures. Furthermore, the addition of extra levels to the switch to increase its throughput, does not cause any noticeable increase in the switching delay.

4. The switch has a regular architecture, which makes it relatively easy to implement in VLSI.

2.6 Conclusion

Several switch architectures have been proposed in recent years for ATM networks [4], [3], [28] etc. These can be broadly classified into time and space division architectures. Due to the high bandwidth requirements of ATM technology, the trend is towards space division architectures which can support multiple connections simultaneously. Recently, a lot of interest has been shown in banyan based architectures due to their several desirable characteristics which suit ATM requirements. In the next chapter we present a new banyan based fast packet switch architecture.

Chapter 3

Proposed Architecture

3.1 Introduction

In this chapter we introduce a novel fast packet switch architecture based on banyan networks called the *parallel-tree banyan switching fabric* (PTBSF). PTBSF is based on a 3-D arrangement of banyan networks in a tree structure. The architecture has been designed to minimize cell loss and latency within the switching fabric. Cell loss can occur only at the lowest level. In the next section we describe its architectural design, the routing algorithm followed by the switch and the manner in which it handles conflict. We then discuss the hardware requirements of the switch. In the end we summarize its features compared to two other known architectures, the tandem banyan [5], and the piled banyan [6].

3.2 Parallel-Tree Banyan Switch

The *parallel-tree banyan switching fabric* (PTBSF) is a fast packet switch architecture employing banyan networks. It is a space division architecture which consists of a 3-D arrangement of banyan networks (or planes) in a *tree-structure* from level 2 onwards as shown in Figure 3.1 for a switch size of $N = 8$. Each level of the tree consisting of 2^{l-2} banyan planes, where l is the level number.

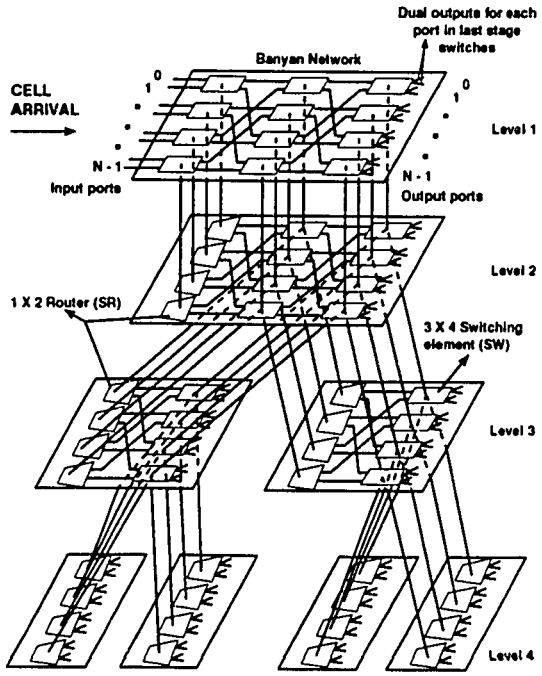

Figure 3.1: The parallel-tree banyan switch architecture (PTBSF).

The basic building blocks within PTBSF are 3-input/4-output switching elements (SE) whose structure is shown in Figure 3.2-a. Each switching element (SW) has two horizontal inputs (I_0 and I_1) and one vertical input (I_v), two horizontal

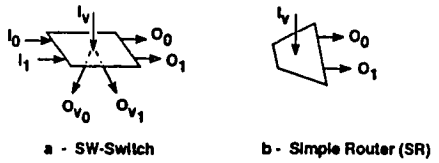

Figure 3.2: Switching elements (SE) in PTBSF.

outputs (O_0 and O_1) and two vertical outputs (O_{v0} and O_{v1}). The packets arrive at the input of the first banyan network in the first level. When a conflict occurs between two packets in the first level, one of the packets is routed correctly, while the other is routed vertically downward to the corresponding SE in the next level.

From level two onwards, there can be three cells at the inputs (I_0, I_1 and I_v) of a given SE requesting the same output. In this case, the switch routes one cell to O_0 or O_1 and the other two cells are routed vertically through the outputs O_{v0} and O_{v1} to the corresponding SEs in the next higher level banyans. This arrangement guarantees conflict-free routing without loss of cells within a SE. For example, in Figure 3.3[1], cells at the inputs 0, 1, 4 and 5, have 0 (000 in binary) as their common destination while cells at the inputs 2, 3 and 6 have 2 (010 in binary) as their destination. Thus, cells at the inputs 0 and 1 compete for output 0 of SE 0, cells at the inputs 2 and 3 compete for output 0 of SE 1 and cells at the inputs 4 and 5 compete for output 0 of SE 2 in the first stage of the banyan at level 1. Suppose there is no conflict at SE 3 of the **first** stage in level 1 and cell 6 passes to the next stage (stage **two**). Let us also assume that cells 0, 2 and 4 are routed to the next

[1]The SEs considered in this example are shaded, while the links are shown with bold lines

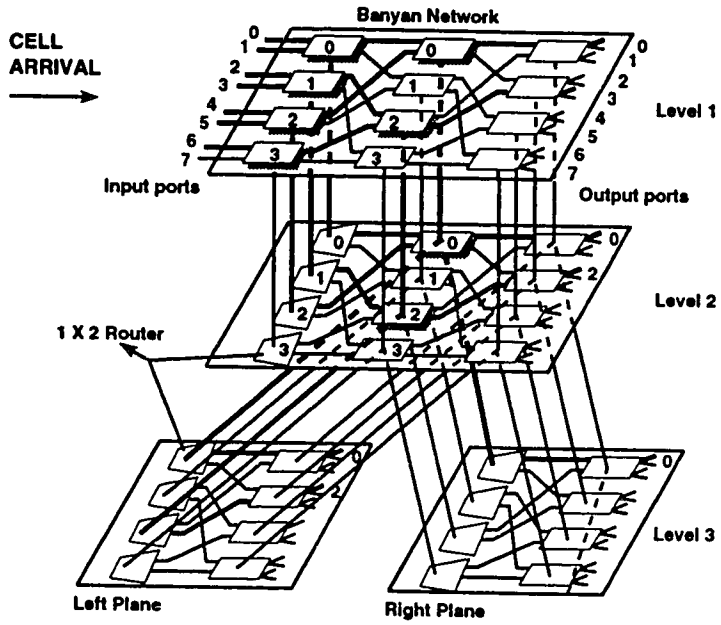

Figure 3.3: Routing in the parallel-tree switch architecture.

stage (stage **two**) in level 1 while cells 1, 3 and 5 are routed vertically downwards to the next level (level 2) of the switch. Cells 0 and 4 contend for output 0 in the SE 0 in stage **two** of level 1 while cells 2 and 6 contend for output 0 in the SE 2 in stage **two** of level 1. Let us assume that cell 0 is routed to the next stage (stage **three**) in level 1 while cell 4 is routed vertically downward to SE 0 in level 2. Also, cell 2 is routed to the next stage (stage **three**) in level 1 while cell 6 is routed vertically downward to SE 2 in level 2.

Thus, SE 0 in the **second** stage of level 2 has three cells (1, 4 and 5) at its inputs. In this situation, cell 1 (say) is routed to the next stage in the same level (level 2) while cells 4 and 5 are routed to the left and right banyan planes in the

next higher level (level 3).

In the second and higher levels, if two cells arrive at a SE destined to the same output, one cell is routed to the correct output while the other is routed to a SE in the same position in the next level banyan. The cell being routed vertically downwards, can be routed either to the left or the right banyan plane in the next higher level of the tree structure. The choice depends on the position of the SE in the plane where conflict has occurred. If the switch position is *even*, the cell is routed to the *left* banyan plane in the next higher level, and to the *right* if the switch position is *odd*. For example, in Fig 3.3, SE 2 in the **second** stage of level 2 has two cells (3 and 6) at its inputs. Since the SE is even numbered, one cell (say cell 3) is routed to the next stage (stage **three**) in level 2 while cell 6 is routed to the left banyan plane. This strategy helps in achieving load balancing in terms of distributing traffic evenly to the banyan planes in the tree structure. The default routing of a cell, to the left or right plane in the next higher level, depending on the SE position, is static and does not require additional hardware.

Thus, the routing algorithm followed by the switch achieves two objectives:

- Cell loss occurs only in the lowest level, and
- Forwarding the cell vertically downward instead of misrouting it, preserves the routing achieved upto that stage. The cell continues routing from the same stage in the next level of banyan.

From second level onwards, there are no cells at the horizontal inputs in the first stage. Hence, there can be no cells forwarded vertically downward at the first stage from level two and above. Thus, from level three onwards, one stage reduces at each level for all the banyan planes in that level. Each banyan network except that in the first level, is made up of two types of SEs. The first stage consists of simple 1-to-2 routers (SR) (see Figure 3.2-b), while the other stages consist of 3-by-4 SEs (SW-switches).

Note that an important feature of PTBSF is that it can achieve a throughput of 100%, i.e., if one is willing to pay the cost in terms of hardware complexity. In Fig 3.1, the last level (level 4) banyans consists of only a single SR stage. It is easy to see that there can be no conflict in any SR stage. Since, cell loss can occur in PTBSF only at the lowermost level, if this level consists of banyans having a single SR stage then there can be no cell loss in PTBSF. The condition under which PTBSF will have only SR stages in the last level banyans is true when the number of levels in a PTBSF is equal to $n + 1$, where $n = log_2 N$ is the number of stages. Thus, PTBSF can conceptually have a 100% throughput.

At each level, the SWs in the last stage of each banyan have the capability to route two packets to the same output as shown by the dual outputs in Figure 3.1. Thus, if two or more cells arrive at a SE in the last stage destined to the same output, two cells can be routed to the correct output in that level itself. This reduces the number of cells traveling to the higher levels and hence reduces cell loss. The outputs

of the banyan networks in all the levels are connected to the corresponding output buffer.

3.3 Hardware resource requirements

A PTBSF with L levels has 2^{L-1} banyan planes. However, the basic element of the first stage for the banyan at level 2 is a simple $1-to-2$ router (SR) (see Figure 3.2-b). In general, each banyan at level k, $k \geq 2$, has one stage that consists of SR switches, and $n-k+1$ stages made of SW's, where $n = log_2 N$ is the number of stages within a single banyan. A SR has much less hardware than an SW. Therefore, we shall be using the number of SW's as a hardware complexity metric of the PTBSF switch. The following Lemma gives the number of SW-stages and SR-stages required for a PTBSF with 2^n inputs and L levels.

Lemma 1 *A PTBSF with $N = 2^n$ inputs and L levels requires $S_{PTBSF}(L) = (n - L + 2)2^{L-1} - 1$ SW-stages and $2^{L-1} - 1$ SR-stages.*

Proof 1 *There is one SR-stage and $n-l+1$ SW-stages in each of the 2^{l-2} banyans of level l, $2 \leq l \leq L$. Therefore, the total number of SR-stages in the PTBSF is $S_{SR}(L) = \Sigma_{l=2}^{L} 2^{l-2}$ that is $2^{L-1} - 1$. The number of SW-stages in the PTBSF is*

$$S_{PTBSF}(L) = n + \Sigma_{l=2}^{L}(n - l + 1)2^{l-2}$$

$S_{SW}(L)$ can be written as $n + n\Sigma_{l=2}^{L} 2^{l-2} - \Sigma_{l=2}^{L}(l-1)2^{l-2}$. Since $\Sigma_{i=1}^{k} i2^{i-1} = (k-1)2^k + 1$, then

$$\Sigma_{l=2}^{L}(l-1)2^{l-2} = (L-2)2^{L-1} + 1$$

The expression of $S_{SW}(L)$ can easily be rewritten as

$$S_{PTBSF}(L) = n + n(2^{L-1}-1) - (L-2)2^{L-1} - 1$$

which simplifies to $S_{SW}(L) = (n - L + 2)2^{L-1} - 1$. ∎

For example, if $n = 3$ and $L = 4$, then the number of banyans in level 3 = $2^{3-2} = 2$. The number of SW-stages in each banyan of level 3 = $(3 - 3 + 1) = 1$. Thus, the total number of SW-stages for a switch of size $N = 2^3 = 8$ with 4 levels is $S_{PTBSF}(4) = (3 - 4 + 2) \cdot 2^{4-1} - 1 = 7$ and the number of SR-stages is $2^{4-1} - 1 = 7$ (see Figure 3.1). One can also make similar statements for the tandem and piled banyan switching fabrics (TBSF and PBSF respectively).

Lemma 2 *A TBSF with $N = 2^n$ inputs and L levels requires nL SW-stages.*

Proof 2 *Obvious from the structure of the TBSF.* ∎

Lemma 3 *A PBSF with $N = 2^n$ inputs and L levels requires $S_{PBSF} = L(2n - L + 1)/2$ SW-stages.*

Proof 3 *In general, the number of SW-stages in level l is equal to $n-l+1$. Therefore, the overall number of SW-stages in all L levels is equal to $\Sigma_{l=1}^{L}(n-l+1)$. which simplifies to $L(2n-L+1)/2$.* ∎

The overall number of SW-switches for a particular switch architecture, i.e., PTBSF, PBSF or TBSF, is equal to the product of the number of SW-stages and the number of SW-switches per stage. We know that for a banyan switch with $N = 2^n$ inputs, the number of elements (SR's or SW's) per stage is equal to 2^{n-1}.

Since the SW-switch dominates the hardware, it is important to evaluate the effective number of $n-$stage banyans constructed with SW-switches only. Such banyans are referred to as SW-banyans.

Definition 1 *The effective number of SW-banyans E_X in a switch of type X, where $X \in \{TBSF, PBSF, PTBSF\}$, having L levels and 2^n inputs is equal to the ratio of the total number of SW-stages S_X in the switch over the number of stages n per banyan that is*

$$E_X(n, L) = \frac{S_X}{n}$$

Using the above definition and lemmas 1 to 3, the effective number of SW-banyans in each of TBSF, PBSF, and PTBSF would be equal to the following:

$$E_{TBSF}(n, L) = L \tag{3.1}$$

$$E_{PBSF}(n, L) = \frac{(2n - L + 1)L}{2n} \tag{3.2}$$

$$E_{PTBSF}(n, L) = \frac{(n - L + 2)2^{L-1} - 1}{n} \tag{3.3}$$

Thus, the effective number of SW-banyans in TBSF is equal to the number of levels. Tables 3.1 and 3.2 summarize the effective number of SW-banyans at each level for PTBSF and PBSF for different switch sizes.

PTBSF						
Levels	Number of SW-Banyans for N =					
(Banyans)	32	64	128	256	512	1024
1 (1)	1	1	1	1	1	1
2 (2)	1.8	1.83	1.86	1.88	1.89	1.9
3 (4)	3	3.16	3.3	3.38	3.44	3.5
4 (8)	4.6	5.16	5.57	5.88	6.1	6.3
5 (16)	6.2	7.83	9	9.88	10.55	11.1
6 (32)	-	10.5	13.57	15.88	17.67	19.1

Table 3.1: Effective number of SW-Banyans in PTBSF.

3.4 Features of PTBSF

In this section we will highlight the most important features of PTBSF in contrast with two other architectures which bear resemblance to our architecture. One of the architecture, proposed by Tobagi et. al. [5], consists of arranging banyan networks

PBSF						
Levels	Number of SW-Banyans for $N =$					
(Banyans)	32	64	128	256	512	1024
1 (1)	1	1	1	1	1	1
2 (2)	1.8	1.83	1.86	1.88	1.89	1.9
3 (3)	2.4	2.5	2.57	2.625	2.67	2.7
4 (4)	2.8	3	3.14	3.25	3.33	3.4
5 (5)	3	3.33	3.57	3.75	3.88	4
6 (6)	-	3.5	3.86	4.125	4.33	4.5

Table 3.2: Effective number of SW-Banyans in PBSF.

in tandem (see Fig 2.11). The other architecture proposed by Hanawa et. al. [6], consists of banyan networks arranged one above the other in a piled structure (see Fig 2.12). The similarities between these two architectures (TBSF and PBSF) and PTBSF are the following.

- They employ banyan networks.
- They have no pre-processing delay.
- They have output buffering.
- They have a multi-outlet design[2].

It is due to these similarities that we will be comparing PTBSF with TBSF and PBSF during performance evaluation under various traffic conditions.

[2]They can transfer multiple packets to the same destination.

Inspite of these similarities, our architecture has several distinctive features. Below, we contrast our architecture (PTBSF) with the tandem (TBSF) and the piled architecture (PBSF).

There are several advantages in arranging banyan networks in a parallel tree structure.

1. In TBSF, once a cell is misrouted at some stage due to a conflict, it has to be routed from first stage onwards in the next banyan in series. Thus, the worst case delay (and hence the switching speed) in TBSF is equal to $L \cdot \tau_b$, where L is the number of levels (banyans) in series and τ_b is the switching time of one banyan.

   In contrast, in PTBSF, when a conflict occurs at some stage, one of the conflicting cells is routed vertically downward to the corresponding switching element in a banyan in the next higher level. Therefore, the vertically forwarded cell continues being routed from the same stage in the next level of banyan. Hence, the cell routing delay and the delay jitter (the difference in the delay suffered by two consecutive cells on a given input, in being routed to the output) are much smaller in the PTBSF than in the TBSF. Actually the (maximum) cell switching delay for the PTBSF will be equal to $\tau + (L-1)$, assuming one clock delay in routing a cell from one level to next.

2. In PTBSF, there is no cell loss except in the higher level contrary to the PBSF where the loss can occur at any level except the first. The throughput of the PBSF reaches some saturation because losses in the upper levels cannot be compensated by adding additional levels. But in PTBSF, adding more levels leads to a considerable improvement in throughput, which is an indicator of good performance scalability as opposed to the saturation effect of the PBSF. The delay performance of PBSF and PTBSF is the same.

3. In the PTBSF there is no pre-processing delay and the cells are switched *on the fly* with least amount of processing delay within the switch. Unlike TBSF, there is no activity and conflict bit processing (refer to Chapter 2), which is required in order to distinguish correctly routed cells from misrouted cells within the SEs as well as at the output of each banyan. The decision to route a cell to a particular output in the PTBSF is based only on the destination bit corresponding to that stage.

3.5 Conclusion

The basic requirements for an ATM switch are high throughput, low latency and in sequence delivery of packets. In this chapter we proposed a novel switch architecture which was designed keeping these objectives in mind. We discussed the way in which cells are routed in PTBSF and its hardware requirements. We also provided a

qualitative comparison between our architecture and the two other related architectures. In the following chapters, we will evaluate the performance of PTBSF both analytically and through computer simulation in order to determine its suitability for ATM. PTBSF will also be compared to two other architectures, namely, TBSF and PTBSF.

Chapter 4

Performance Evaluation under Uniform Traffic

4.1 Introduction

Any new design has to be validated by being tested thoroughly under the conditions for which it has been designed. There are several methods in order to evaluate the merits of an ATM switch architecture. One of the most widely used methods is to analyze the performance of the switch under uniform traffic. A uniform traffic is defined as one in which the cell arrival at an input follows a uniform distribution pattern. The destinations (output ports) of the cells may be uniformly distributed, i.e., all the output ports have an equal probability of being selected or may be selected according to some fixed pattern (as in *communities of interest*

traffic, discussed later).

Although the uniform traffic model is not an accurate representation of actual ATM traffic, it is often used as a common testbed in order to compare different switch architectures. The popularity of the uniform model stems from the fact that it is mathematically tractable. In this chapter we present an analytical model, based on uniform traffic, for PTBSF, TBSF and PBSF. We will also present the results of the performance evaluation of the above three architectures as obtained through computer simulations under uniform traffic for the following three cases:

- destinations of the cells are selected uniformly.
- destinations of the cells are selected according to a fixed pattern (as in "communities of interest" traffic).
- destinations selected by the cells are a permutation of the set of output ports.

4.2 Uniform Traffic With Destinations Randomly Selected

4.2.1 Analytical Model

We derived the analytical expressions for the cell loss probability of the PTBSF, PBSF and TBSF, assuming uniform traffic at the switch inputs. We refer the reader to the switches shown in Figure 3.2.

For the case of PTBSF, assume that a cell arrives at either input I_0 or I_1 of an SW-switch with a probability p, and that the probability of a cell arrival on the vertical input I_v is q. Since the SW-switch has three inputs, we analyze the following cases:

1. There is only one cell at any of the horizontal input I_0 or I_1 and no cell at the vertical input I_v with probability $p(1-p)(1-q)$ or there is a cell at the vertical input I_v and no cell at the horizontal inputs I_0 and I_1 with probability $(1-p)^2 q$ or

2. There are two cells at I_0 and I_1 and no cell at the vertical input I_v with probability $p^2(1-q)$ or there are two cells at I_0 (or I_1) and I_v with no cell at I_1 (or I_0) with the probability $p(1-p)q$.

3. There are three cells at the inputs I_0, I_1 and I_v with probability $p^2 q$.

The probability that a cell exits the switch at a given output (pass-through) among O_0, O_1, O_{v0}, and O_{v1} is shown in Table 4.1 for an arbitrary SW-switch. For a simple router (Figure 3.2-b), if a cell arrives at the input I_v with probability q, the probability that the cell exits at either outputs O_0 or O_1 is $q/2$. This model allows propagating the pass-through probability in the horizontal direction and finding the probability of the cell being routed to the next higher level through the vertical outputs. The loss probability in the PTBSF is the sum of all the probabilities of losses occurring at the outputs O_{v0} and O_{v1} of all the SW-switches located in the

Switch	Cells	$P(O_0)$ or $P(O_1)$	$P(O_{VO})$	$P(O_{V1})$	P(loss)
PTBSF	One cell	$(1-p)(1-q)p+(1-p)^2q/2$	0	0	0
	Two cells	$3(1-q)p^2/4+3(1-p)pq/4$	$(1-q)p^2/2+(1-p)pq$	0	0
	Three cells	$7p^2q/8$	p^2q	$p^2q/4$	0
TBSF	One cell	$p(1-p)$	-	-	0
	Two cells	$3p^2/4$	-	-	$p^2/2$
PBSF	One cell	$(1-p)(1-q)p+(1-p)^2q$	0	0	0
	Two cells	$3(1-q)p^2/4+2(1-p)pq$	$(1-q)p^2/4+(1-p)pq$	$P(O_{VO})$	0
	Three cells	p^2q	$3p^2q/4$	$3p^2q/4$	$p^2q/4$

Table 4.1: Analytical model under uniform traffic of the PTBSF

highest level of the PTBSF.

The cell loss probabilities of TBSF and PBSF have been similarly derived and are summarized in Table 4.1.

4.2.2 Computer Simulation

In this section we present the cell loss performance of PTBSF, TBSF, and PBSF obtained through computer simulation. We assume a time slotted synchronous operation of the switch, where the *slot size* is equal to the switch processing time. Cells must be synchronized and aligned with the local slot boundaries before being routed by the switch. The workload assumptions are based on a uniform traffic model and are as follows.

- All switch inputs are identical and independent.
- In every time slot, each input in the first stage of the first level has a probability p of having a cell and $1-p$ of having no cell.

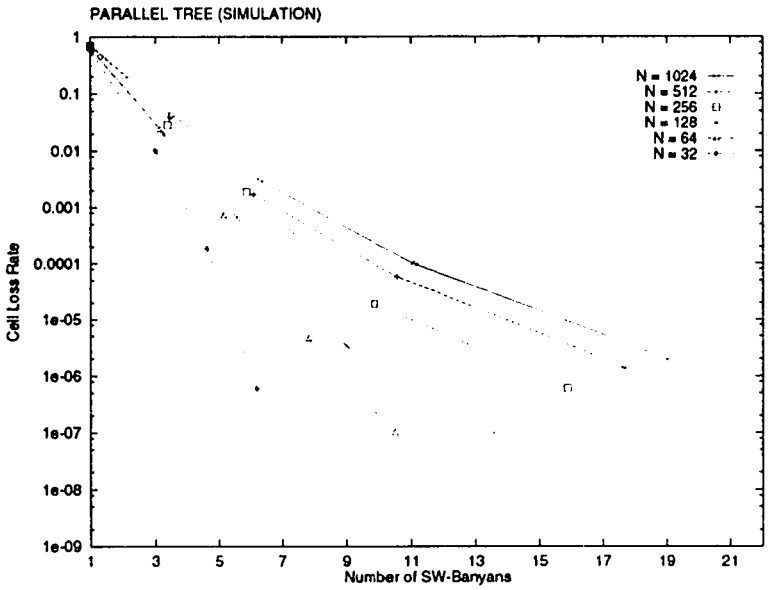

Figure 4.1: Loss rate in PTBSF at $p = 1$ for different switch sizes (Simulation).

- We also assume that activity level at the various inputs are independent in time as well as in space.

- In each time slot, the cell destinations are randomly distributed over all the switch outputs according to a uniform distribution.

We built simulators to simulate the operation of PTBSF, TBSF and PBSF. The three architectures were then subjected to uniform traffic under different loads. The results were obtained using an *omega* topology for the banyan networks.

Figure 4.1 shows the cell loss rate in PTBSF versus the number of SW-banyans for various values of N at full load ($p = 1$) obtained through simulation. Most ATM traffic sources require a cell loss rate lower than 10^{-6} [5]. From the figure it is clear that a loss rate of 10^{-6} is achievable for all the considered ($32 \leq N \leq 1024$) switch

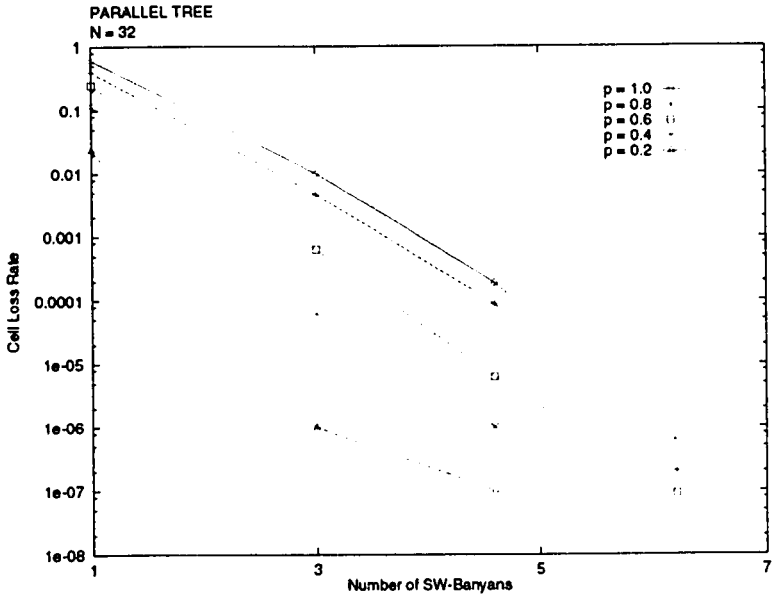

Figure 4.2: Loss rate in PTBSF at different loads for $N = 32$ (Simulation).

sizes. To obtain a cell loss rate of 10^{-6}, the number of SW-banyans should be 6.2 (4-levels) for $N = 32$ and 19.1 (5-levels) for $N = 1024$.

Figures 4.2 and 4.3 show the cell loss rate obtained by varying the load p for $N = 32$ and $N = 1024$ respectively. As we decrease the load, the number of banyans required to achieve a given packet loss rate decreases. This decrease is more prominent as the switch size becomes larger ($N = 1024$). In PTBSF, the banyans in each subsequent layer are subjected to decreasing loads (only the conflicting cells go down to the next level). Thus, the above result is useful in order to predict the throughput of a banyan in a given layer of PTBSF.

Fig 4.4 shows the percentage of cells lost versus the stage number in PTBSF without dual outputs (actual PTBSF has dual outputs for each port in the last stage SEs) for $N = 256$ at full load. As seen from the figure, there is no loss at

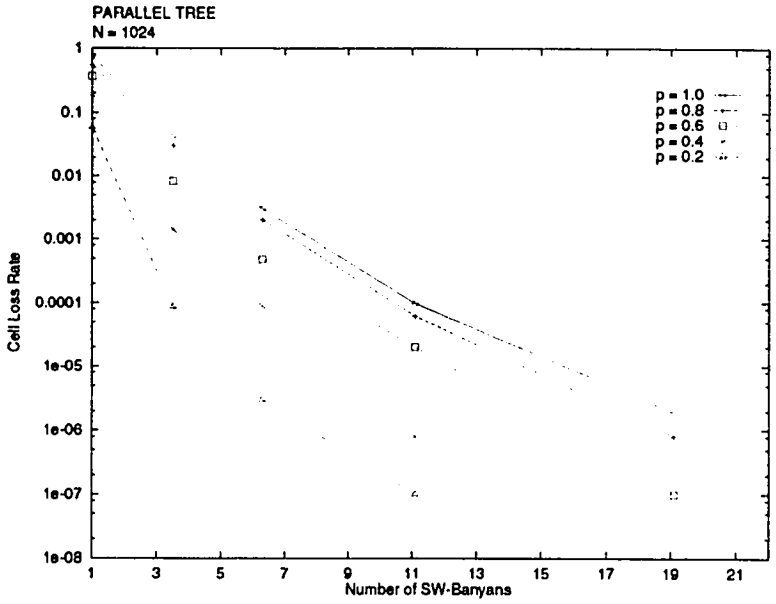

Figure 4.3: Loss rate in PTBSF at different loads for $N = 1024$ (Simulation).

any of the first three stages of the SW-banyans at the lowest level. From fourth stage onwards, percentage of cell loss increases at each subsequent stage, with the loss being maximum in the last stage. This lead us to have dual outputs (for each output port) in the SW-switches of the last stage in each banyan plane. Dual outputs in the last stage banyans minimize the number of cells being routed to the lower layers.

Fig 4.5 shows the cell loss ratio between levels i and $i+1$ versus the level number in PTBSF for various switch sizes at full load. The cell loss decreases at a faster rate as the switch size becomes smaller. This is expected since the likelihood of conflicts increases as the switch size becomes larger.

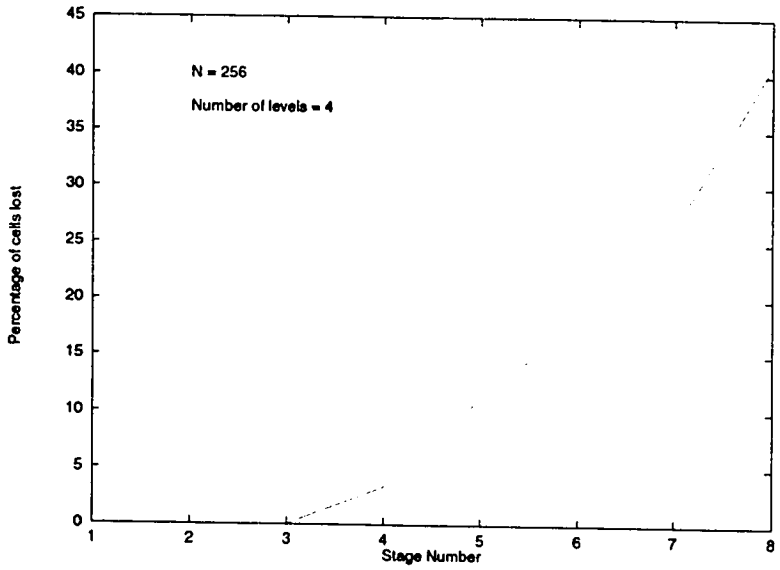

Figure 4.4: Percentage of cells lost in PTBSF at various stages for $N = 256$ at full load.

4.2.3 Comparison between the analytical and simulation results

Figure 4.6 shows the cell loss rate performance obtained from the analytical model of the PTBSF as a function of the number of SW-banyans and for various values of N at full load ($p = 1$). The analytical model matched well with the simulation results as shown in Figure 4.7, especially when the number of levels is relatively small. This validates the obtained results. The cell loss rates obtained from the analytical model were slightly optimistic (smaller) than those obtained from simulation. The deviation between the analytical and simulation results slightly increases with the number of levels. One reason for this deviation is that the random number generator used is not ideal. Moreover, the reason the deviation increases with the number of

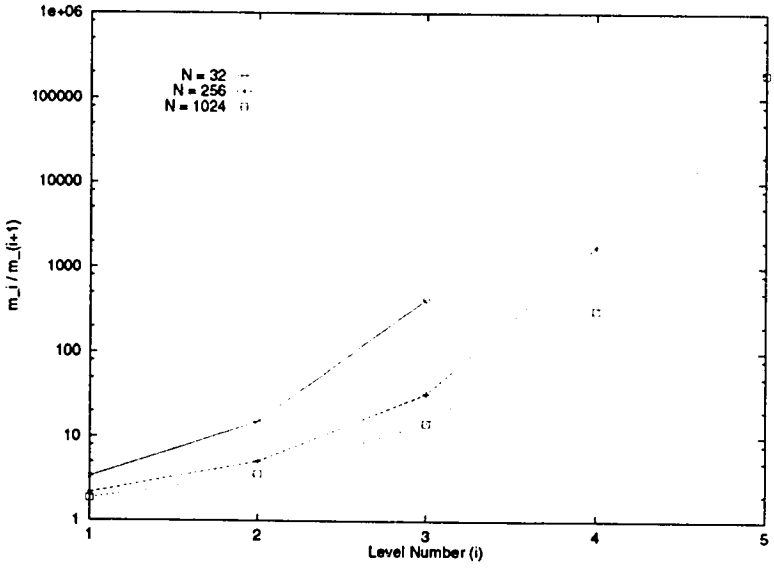

Figure 4.5: Ratio of cells lost in PTBSF at various levels for different switch sizes at full load.

levels is that, the independence assumption between the various traffic streams is not preserved from the second level onwards. The cells forwarded to the lower levels are somehow correlated in space. The degree of correlation increases from one level to the next. Correlation in space or time results in increased cell loss. Therefore, since the analytical model assumes cell independence across all levels, the analytical results are more optimistic, and for that matter less accurate than the simulation results.

4.2.4 Comparing PTBSF to PBSF

In this section, we will compare PTBSF to another parallel architecture, the PBSF. It also consists of arranging banyans in a 3-D structure.

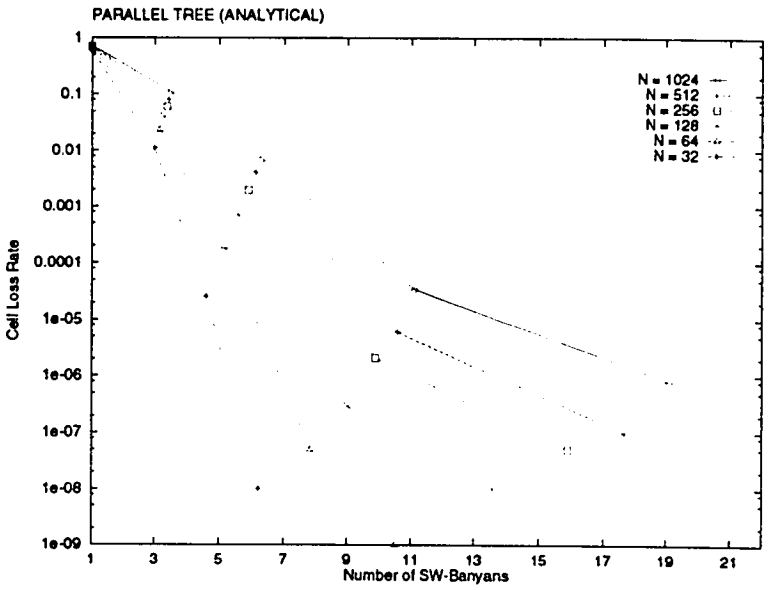

Figure 4.6: Loss rate in PTBSF at $p = 1$ (Analytical).

An important feature of PTBSF is its capability to achieve a high throughput with efficient utilization of resources. In PTBSF, the cell loss rate sharply decreases when we increase the hardware resources (more levels are added).

In PBSF it is not so. Figures 4.8 and 4.9 show the cell loss rate in PBSF versus the number of SW-banyans for different switch sizes using the simulation and the analytical models respectively. As seen from the figures, the throughput of PBSF saturates at some value for all switch sizes, no matter how many levels are added.

Let m_i and m_{i+1} be the cell loss rates at levels i and $i+1$ respectively. Figure 4.12 shows the variation of the ratio m_i/m_{i+1} versus the number of SW-banyans, for the TBSF, PBSF, and PTBSF. Figure 4.12 clearly indicates that cell loss decreases at a much faster rate with each additional level for PTBSF than for PBSF.

In fact, each additional banyan level decreases the cell loss rate by at least an

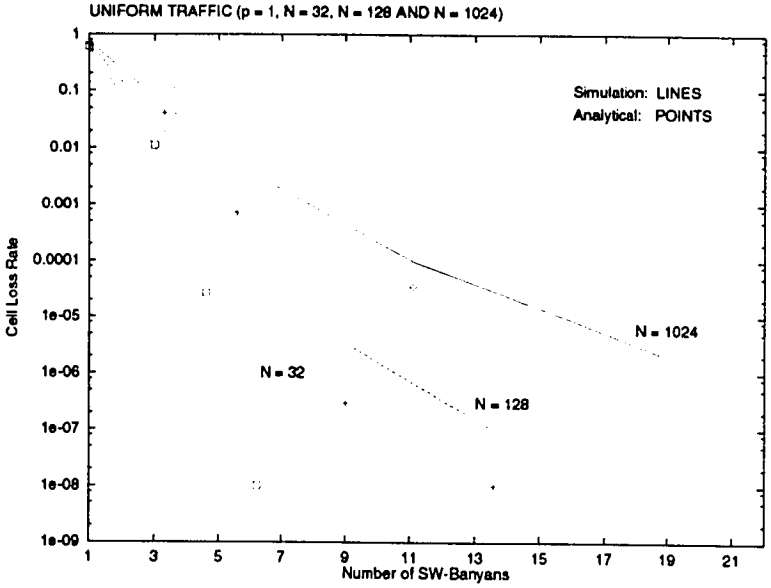

Figure 4.7: Loss rate in PTBSF at $p = 1$ for different N.

order of magnitude. From Figure 4.1 and 4.8, for $N = 256$, the cell loss rate decreases from 10^{-3} to approximately 10^{-5} when the number of SW-banyans increases from 5.88 (level 3) to 9.88 (level 4). In contrast, in PBSF, for $N = 256$, the cell loss decreases from 10^{-1} to less than 10^{-2} when the number of SW-banyans increases from 2.6 (level 3) to 3.25 (level 4). Although in PTBSF, the amount of hardware resources increased with each level is significantly higher than in PBSF, it pays off in terms of considerably higher throughput as is required for ATM.

Moreover, the cell loss rate ratio (m_i/m_{i+1}) decreases in case of PBSF in going from level 4 to level 5. This is contrary to what one might expect, because the SW-banyans are supposed to have higher throughput at lower levels due to much lower load. The reason for this is that in PBSF, loss can occur anywhere within the

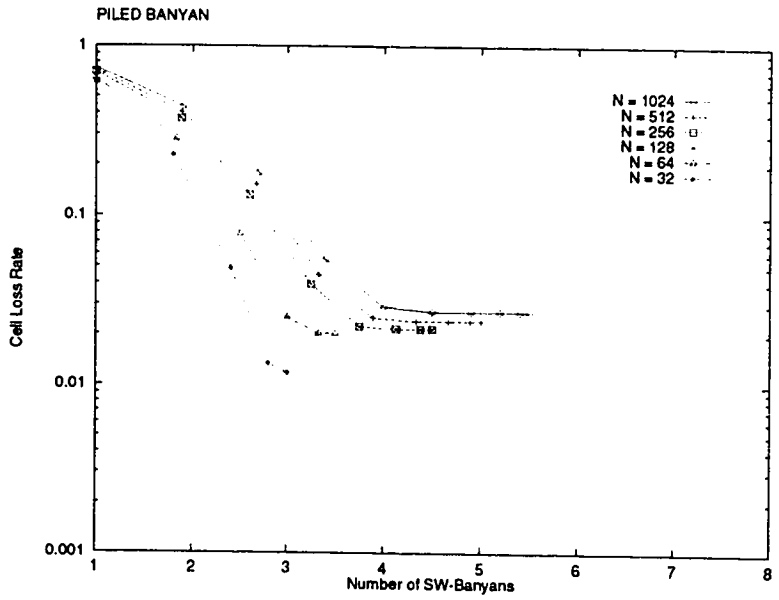

Figure 4.8: Loss rate in PBSF at $p = 1$ for different switch sizes (Simulation).

switch except in the first level [1]. Thus, the additional hardware resources cannot compensate for the losses which have occurred in the upper layers. Therefore, increasing the hardware resources does not guarantee increase in the throughput in case of PBSF. In contrast, in PTBSF, cell loss can occur only at the lowest level, resulting in guaranteed increase in throughput, with each additional level.

This proves that the PTBSF is one *parallel architecture* that overcomes the saturation problem which severely limits the throughput of PBSF. In PBSF, the cell loss rate remains constant at about 10^{-2} even if we increase the number of levels beyond 4. Saturation occurs in PBSF even when the load is as low as $p = 0.2$. For the PBSF, the saturation effect occurs at a loss rate of about 10^{-2} for $p = 1$ and at

[1]Note that a cell loss can occur at any level in PBSF when three cells compete for one output of some switching element.

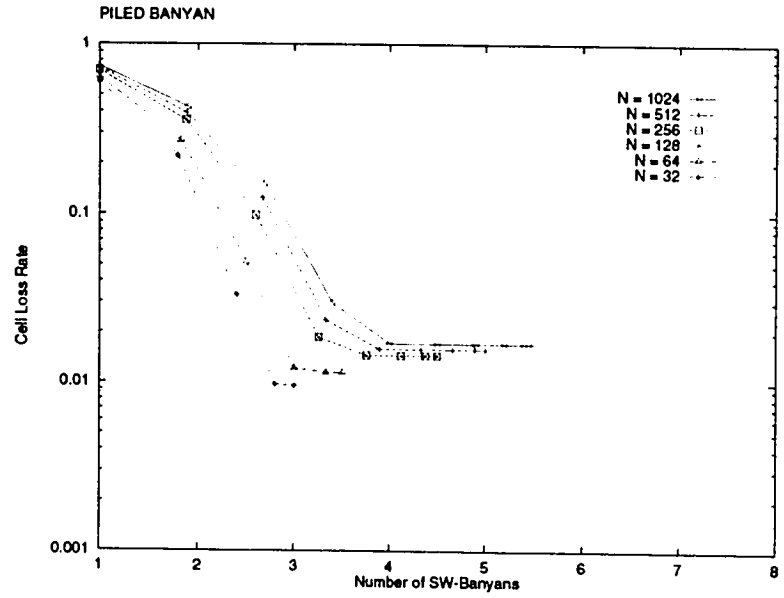

Figure 4.9: Loss rate in PBSF at $p = 1$ for different switch sizes (Analytical).

about 10^{-4} for $p = 0.2$ for switch size $N = 256$ [6]. The saturation effect limits the performance scalability of switches such as the PBSF, since the use of additional levels does not guarantee an increase in the throughput.

4.2.5 Comparing PTBSF to TBSF

In this section we will compare PTBSF to another architecture which consists of arranging banyans in series, the TBSF [5]. If we recall from Chapter 2, in TBSF, when a conflict occurs within a switching element one cell is correctly routed and the other is misrouted. To distinguish misrouted cells from the correctly routed ones, an *activity bit* and a *conflict bit* is appended to the cell header. At the output of each banyan, correctly routed cells are removed and stored into their corresponding output buffers, and misrouted cells are fed to the next banyan for further processing.

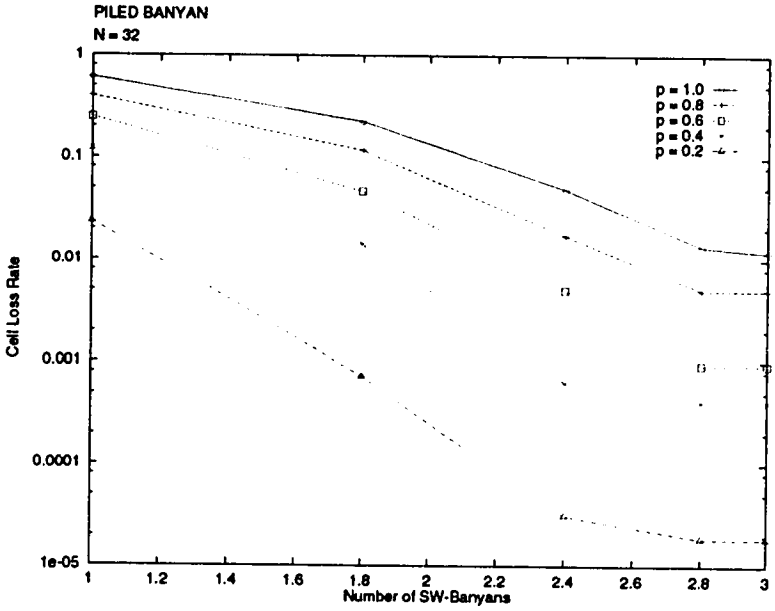

Figure 4.10: Loss rate in PBSF for $N = 32$ at different loads (Simulation).

We evaluated the performance of TBSF under uniform traffic pattern through simulation for N ranging from 32 to 1024. Figure 4.13 shows the cell loss rate in TBSF for different switch sizes at full load. As seen from the figure, in order to achieve a cell loss rate of 10^{-6}, 9 banyans are required for $N = 32$ and 14 banyans are required for $N = 1024$. Figures 4.14 and 4.15 show the cell loss rate in TBSF for $N = 32$ and $N = 1024$ respectively at different loads. Both TBSF and PTBSF exhibit acceptable cell loss rates (10^{-6} or better) as required by most ATM traffic sources (See Figures 4.1 and 4.13).

However in TBSF, cell processing is more complex because each switching element has to process a conflict bit and an activity bit in addition to the destination bit. More processing is required in order to distinguish correctly routed cells from the misrouted ones at the output of each banyan network. This constitutes an

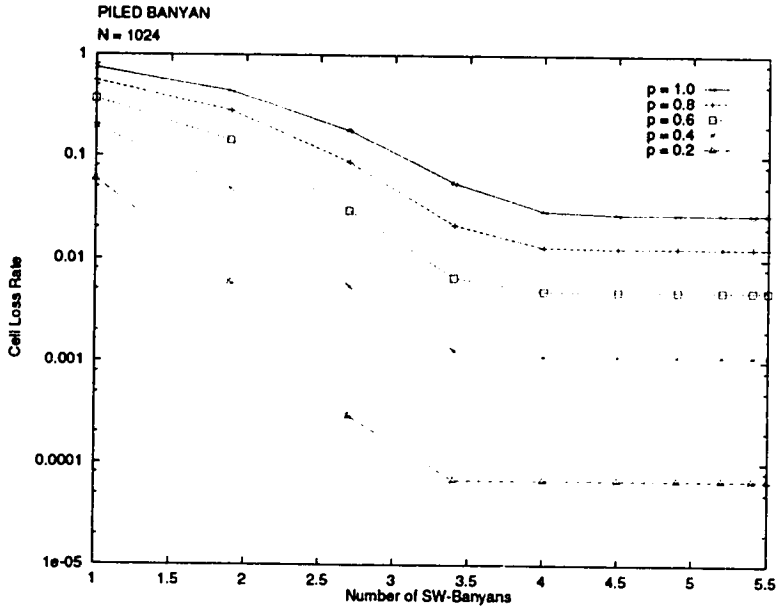

Figure 4.11: Loss rate in PBSF for $N = 1024$ at different loads (Simulation).

overhead, both in terms of latency and hardware cost.

In PTBSF, although each individual SW-switch is more complex, the total number of SW-banyans required in order to achieve a given cell loss rate (10^{-6}) is lower than TBSF, for switches of sizes upto $N = 128$. Figures 4.16 to 4.21 show the cell loss rate versus the number of SW-banyans required for PTBSF, TBSF and PBSF at full load for various switch sizes. In each of these figures, the cell loss rate for PBSF saturates at about 10^{-2}. The number of SW-banyans required to achieve a cell loss rate of 10^{-6} in TBSF is higher than PTBSF upto switches of size $N = 128$. For $N = 256$, upto a cell loss of 10^{-5}, the number of SW-banyans required by PTBSF is lower than TBSF. When the number of SW-banyans is about 10 and the cell loss rate is 10^{-5}, TBSF overtakes PTBSF in terms of the number of SW-banyans required to achieve the same cell loss performance. We refer to this as the *crossover-point.*

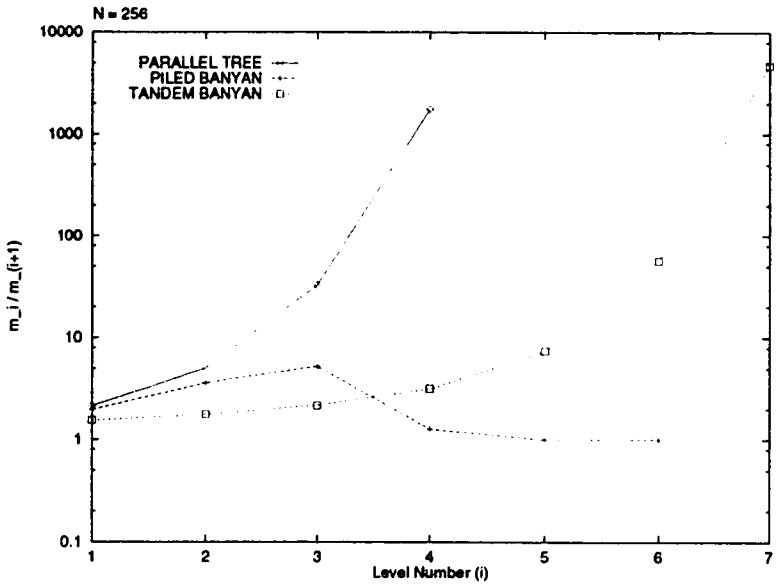

Figure 4.12: Loss ratio at different levels in the three architectures for $N = 256$ at full load.

Similar crossover-points occurs in the Figures for $N = 512$ and $N = 1024$, although at a lower cell loss rate (10^{-4}). This goes on to show the feasibility and superiority of PTBSF, especially for switches of smaller sizes.

Two of the main problems of TBSF are the potential of out-of-sequence cell delivery and the large delay jitter that is due to its sequential structure. For the TBSF, in each time slot the variance in the routing time of cells to the output buffers can be as large as $(K-1)\tau_b$, where K is the number of banyans and τ_b is the propagation time for one banyan. Since the order of cells in the FIFO buffers depends on their arrival times, then there is a possibility for cells reaching the buffer in consecutive time slots to be out-of-sequence.

To avoid this problem, cells reaching the buffer from each banyan are delayed so

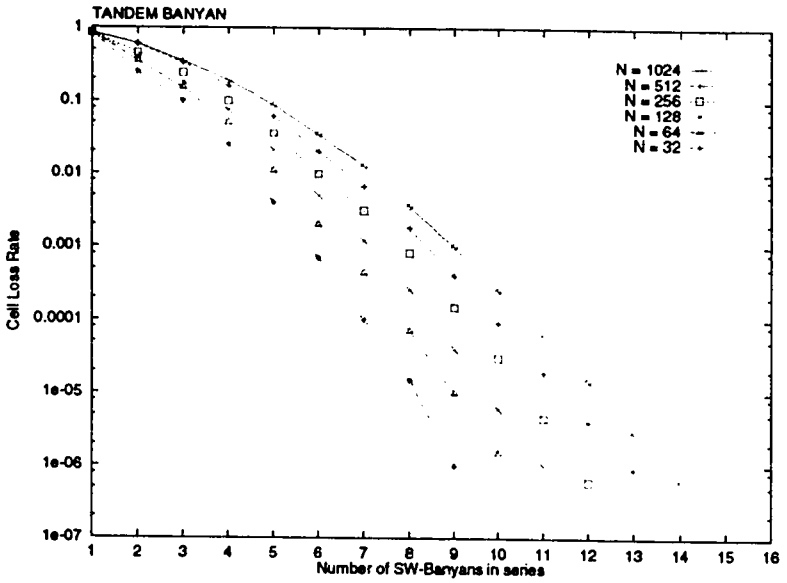

Figure 4.13: Loss rate in TBSF at $p = 1$ for different switch sizes (Simulation).

that their arrival time to the buffer becomes $K \cdot \tau_b$ regardless of their routing. To keep the loss rate below 10^{-6} under full load, the value of K should be at least 14, which means that the TBSF delay is 14 times the propagation delay of one banyan switch. Though the switching time may be just a fraction of the total routing time of the cells, the switching delay of the TBSF is much slower compared to that of the PTBSF. Moreover, due to the random delay in routing the cells at a given input to the switch output[2], the *variance of delay* between the cells is likely to be high.

In the PTBSF, a cell reaches the output in time τ_b at the earliest and in time $\tau_b + (L-1)$ at the latest, where L is the number of levels. Furthermore, since the value of L is below 6 in all the studied PTBSF configurations, the delay difference

[2]Cell delay can be anywhere between τ_b (if the cell exits the first banyan) to $K \cdot \tau_b$ (if the cell exits the last banyan).

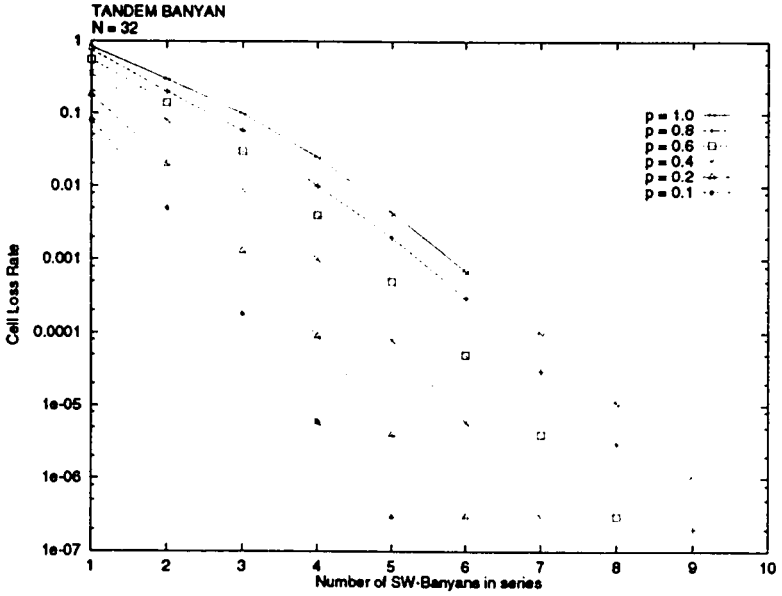

Figure 4.14: Loss rate in TBSF for $N = 32$ at different loads (Simulation).

between a cell exiting at first level and a cell exiting the switch at level L is quite small. Figure 4.22 shows the delay performance of TBSF compared to PTBSF. For TBSF, the delay rises linearly as a function of the switch size. On the other hand in PTBSF, the rise in delay with the switch size is almost negligible. In fact, the delay in the TBSF is 5 to 8 times the delay in the PTBSF for switch sizes ranging from $N = 32$ to $N = 1024$.

4.3 Communities of Interest Traffic

Sometimes the traffic pattern encountered by an ATM switch may be such that the contention between the cells is mainly due to internal congestion rather than output conflicts. This corresponds to the "*communities of interest*" traffic in which all cells

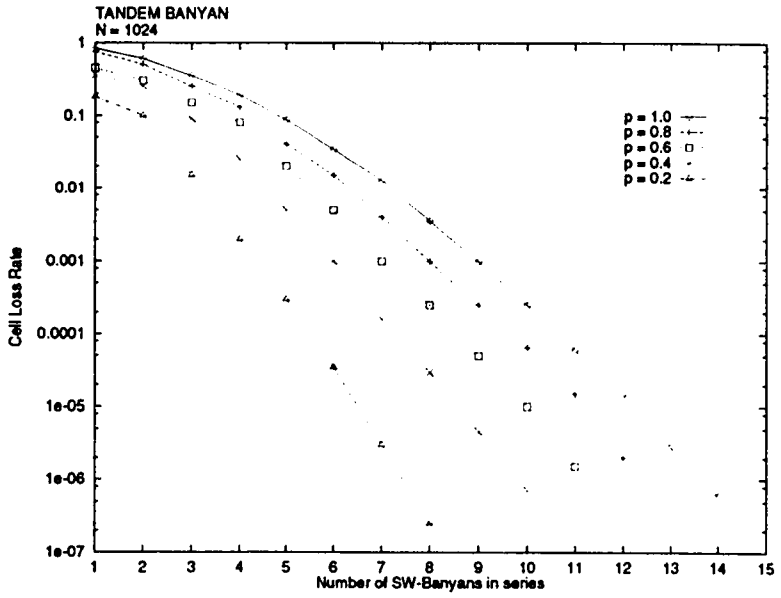

Figure 4.15: Loss rate in TBSF for $N = 1024$ at different loads (Simulation).

entering the first M inputs (for example) are destined to the first M outputs, all cells entering the second group of M inputs are destined to the second group of M outputs and so on as shown in Figure 4.23.

In this kind of traffic scenario, the input-output pairs which cause the internal conflict due to their paths sharing the same interconnection links in the middle of the network constitute the "*communities*". This traffic pattern has been studied in [5]. In this section we will present the cell loss behavior in the three architectures under this kind of traffic.

We simulated the performance of the three architectures under the communities of interest type of traffic. We assume the group size to be four. The results are shown in Figures 4.24 and 4.25 for switch size $N = 32$ and $N = 64$ respectively. As can be seen from the figures, the cell loss rate saturates for PBSF at about 0.25

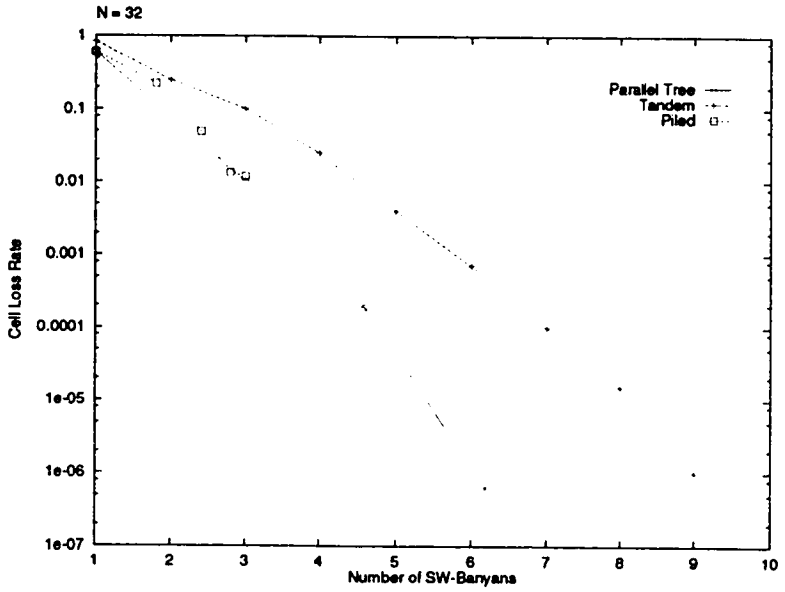

Figure 4.16: Cell Loss rate in PTBSF, TBSF and PBSF under uniform traffic for $N = 32$ at full load (Simulation).

for $N = 32$ and at 0.35 for $N = 64$. The performance of PBSF is particularly inferior because the communities of interest traffic pattern causes several instances of three cells arriving at a switching element inputs, requesting the same output. This leads to a noticeable increase in cell loss in PBSF. In contrast, PTBSF gives an excellent performance under this scenario, because it can very efficiently route all three arriving cells at a switching element input. Thus, in PTBSF, a cell loss rate of 10^{-6} is achieved with 3 SW-banyans (3 levels) for $N = 32$ and with 5.16 SW-banyans (4 levels) for $N = 64$. For TBSF, 8 SW-banyans are required for $N = 32$ and 9 SW-banyans are required for $N = 64$. This shows the superiority of PTBSF in comparison to TBSF and PBSF under this type of traffic.

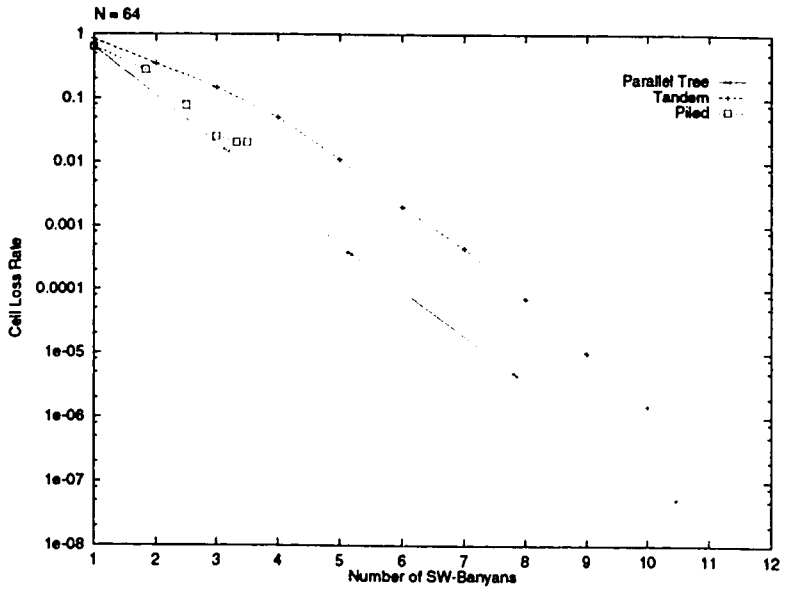

Figure 4.17: Cell Loss rate in PTBSF, TBSF and PBSF under uniform traffic for $N = 64$ at full load (Simulation).

4.4 Permutation Traffic

This kind of traffic is generally encountered in *synchronous transfer mode* (i.e., circuit switching) and has been studied by Tobagi et. al. [5]. In *permutation traffic*, no two cell destination requests, in a given time slot are the same. Thus, loss occurs only due to internal blocking and not due to output conflicts.

Although PTBSF has been designed for ATM networks, we compared the performance of the three architectures under this kind of traffic too. Figures 4.26 and 4.27 show the cell loss rate under permutation traffic for $N = 32$ and $N = 64$, under full load.

As seen from the figures, PBSF saturates after reaching a certain cell loss rate (0.0025 for $N = 32$ and 0.005 for $N = 64$ respectively). PTBSF performs better

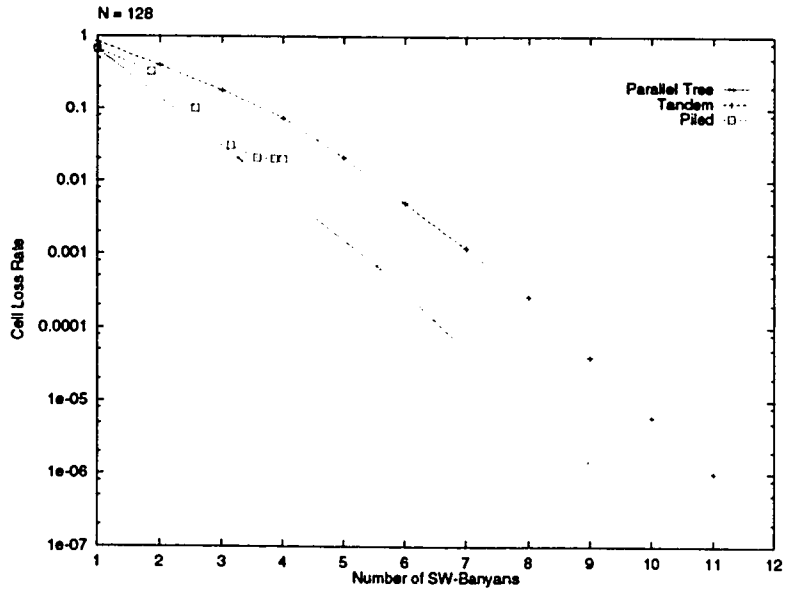

Figure 4.18: Cell Loss rate in PTBSF, TBSF and PBSF under uniform traffic for $N = 128$ at full load (Simulation).

than TBSF for a switch size of $N = 32$. But, for $N = 64$, TBSF overtakes PTBSF in performance . This suggests that arranging banyan networks in tandem may be a better solution to tackle internal blocking rather than arranging them in parallel.

4.5 Conclusion

This chapter discussed the performance evaluation of the three architectures, namely PTBSF, TBSF and PBSF under independent uniform traffic pattern. Though, this pattern is not encountered in real ATM networks, it provides a common platform in order to compare different architectures. PTBSF performs reasonably well, both in terms of throughput, as well as latency. The results presented in this chapter validates our proposed design, in addition to emphasizing the merits of arranging

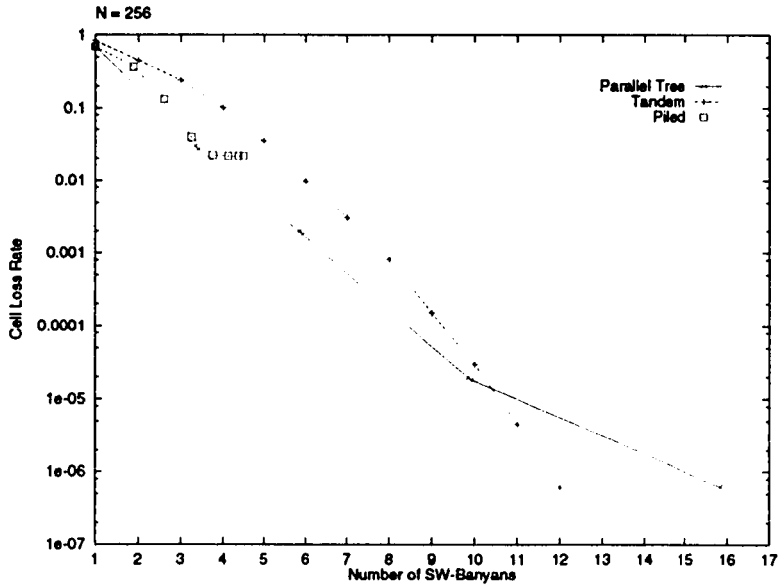

Figure 4.19: Cell Loss rate in PTBSF, TBSF and PBSF under uniform traffic for $N = 256$ at full load (Simulation).

banyan networks in a tree structure. In the next chapter, we present the assessment results of the three architectures when their simulation models were subjected to *realistic* ATM traffic.

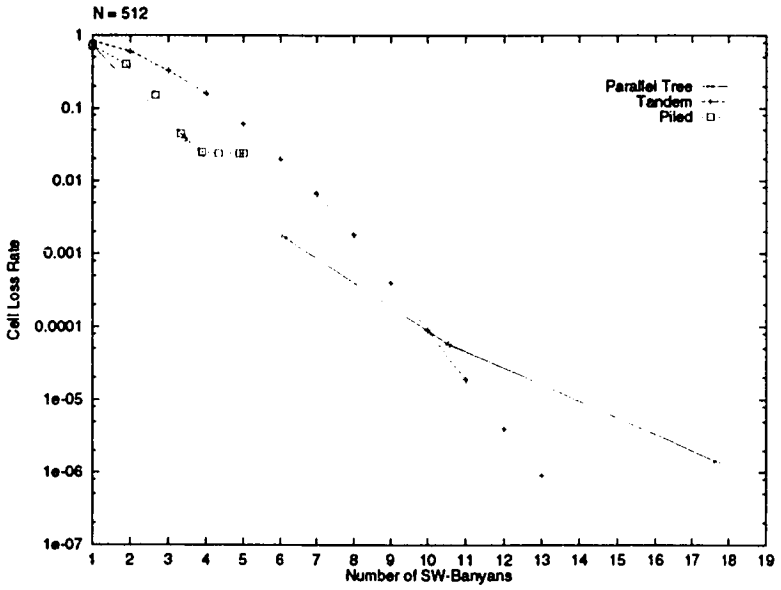

Figure 4.20: Cell Loss rate in PTBSF, TBSF and PBSF under uniform traffic for $N = 512$ at full load (Simulation).

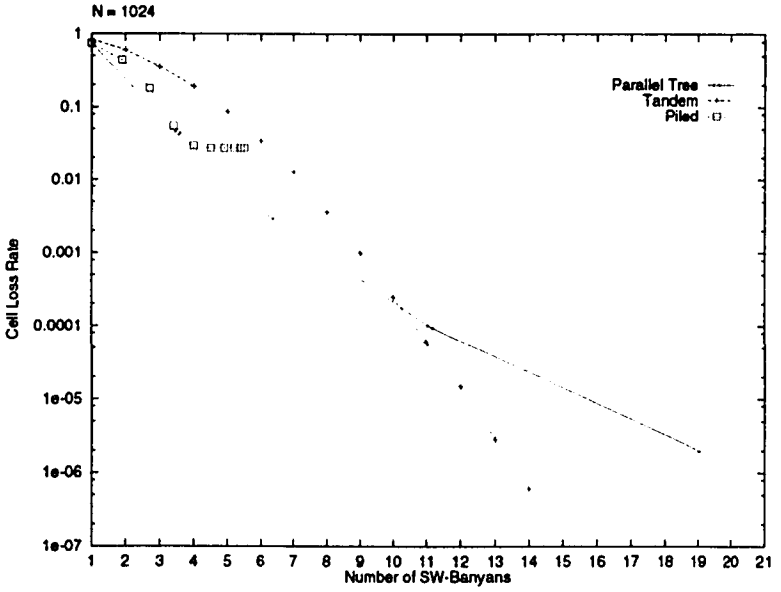

Figure 4.21: Cell Loss rate in PTBSF, TBSF and PBSF under uniform traffic for $N = 1024$ at full load (Simulation).

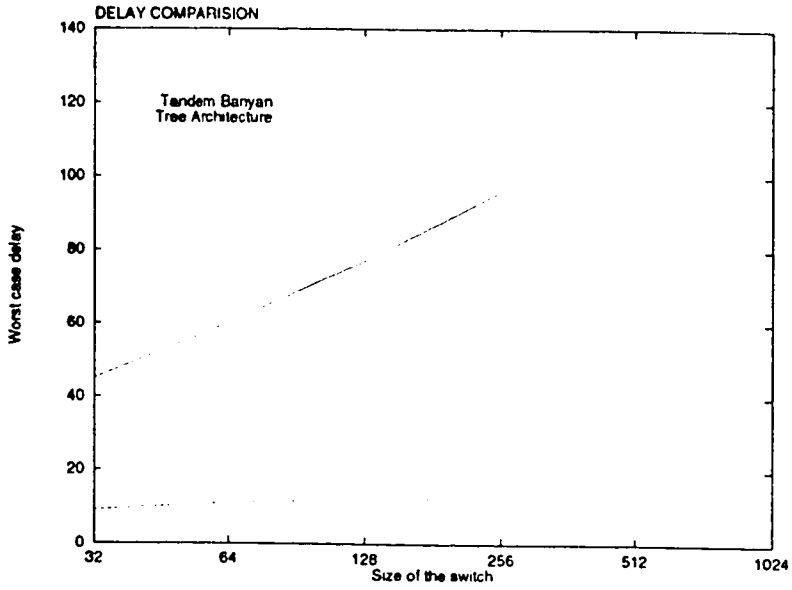

Figure 4.22: Worst case delay in TBSF and PTBSF for different switch sizes.

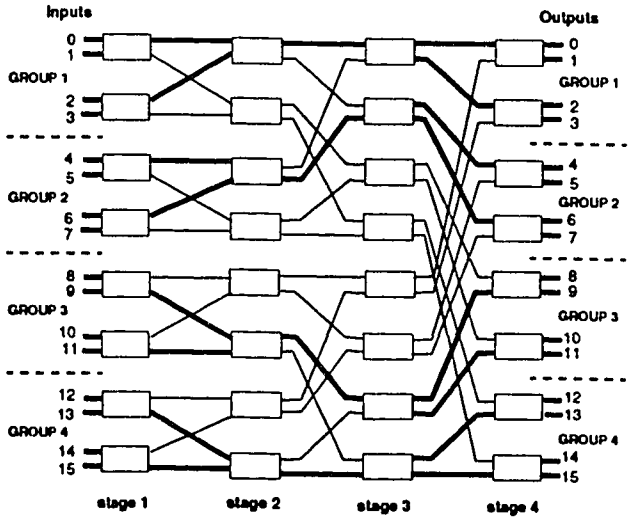

Figure 4.23: Maximum internal congestion in a 16 × 16 Banyan.

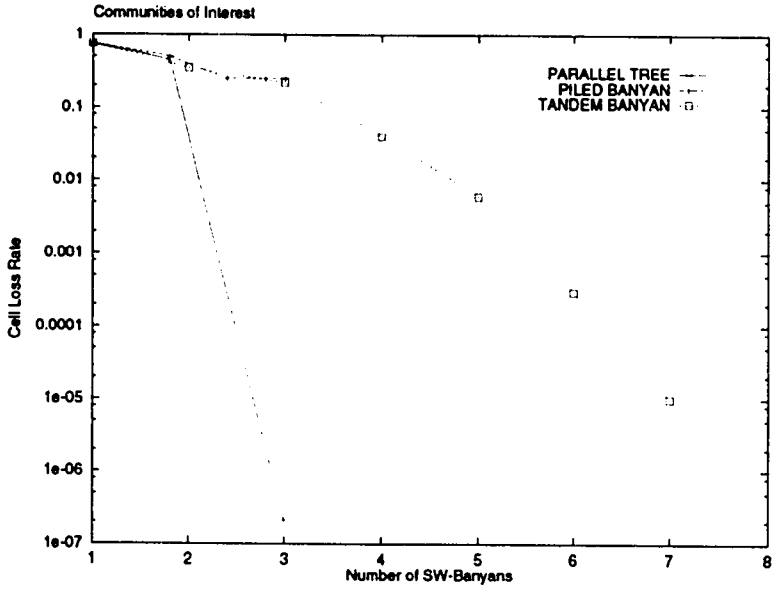

Figure 4.24: Cell loss performance of TBSF, PBSF, and PTBSF when subjected to a *communities of interest traffic* for $N = 32$ at $p = 1$.

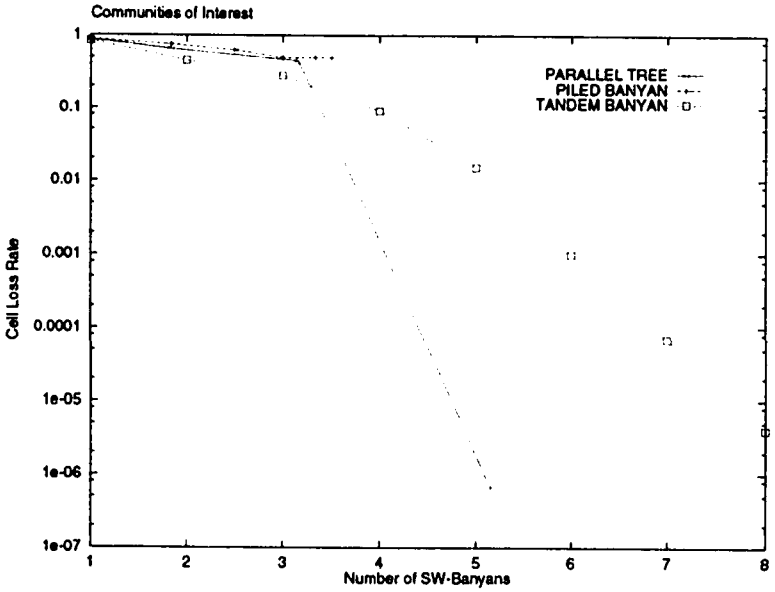

Figure 4.25: Cell loss performance of TBSF, PBSF, and PTBSF when subjected to a *communities of interest traffic* for $N = 64$ at $p = 1$.

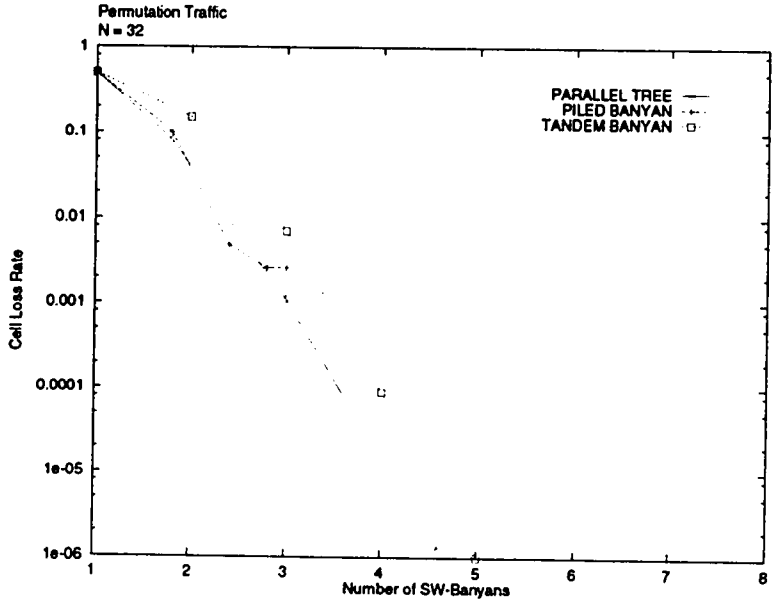

Figure 4.26: Cell loss performance of TBSF, PBSF, and PTBSF when subjected to a *permutation traffic* for $N = 32$. We assume full load at the inputs.

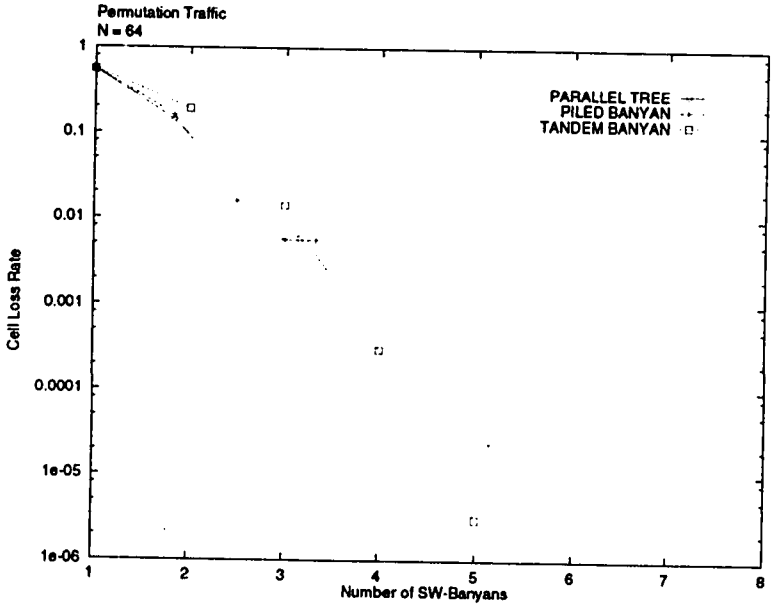

Figure 4.27: Cell loss performance of TBSF, PBSF, and PTBSF when subjected to a *permutation traffic* for $N = 64$. We assume full load at the inputs.

Chapter 5

Performance Evaluation under ATM Traffic Conditions

5.1 Introduction

ATM networks are expected to support a wide range of traffic sources. There is a general agreement that the various traffic sources to be serviced by ATM networks belong to the following three main classes:

1. Variable Bit Rate (VBR) sources. These mainly include computer related data sources, such as file transfer, electronic mail, or terminal emulation.

2. Constant Bit Rate (CBR) sources. These consists mainly of applications such as telephony, voice mail, or sometimes teleconferencing.

3. VBR-Video sources. Sources of this type are mainly the result of multimedia applications.

In order to determine whether a particular switch architecture is suitable for broadband ATM or not, one has to observe the performance of the switch when subjected to the type of traffic expected in ATM networks. A realistic traffic mix will subject the switch to a genuine workload. A realistic workload has several desirable characteristics from testing point of view, namely,

1. Traffic load will be from a variety of sources, bursty and nonbursty.

2. The traffic sources will exhibit correlation in space as well as in time. This correlation will be due to several factors such as, burstiness of the sources, and the pairing of input ports with the output ports. Such correlation will last for the entire duration of the connections, not just for one burst.

In this chapter, we will first discuss the traffic source models and their usefulness with a detailed description of the *ON-OFF traffic source model.* In the next section, we will discuss the performance of PTBSF, TBSF and PBSF when subjected to a realistic ATM workload consisting of the above mentioned traffic sources. We used the *ON-OFF traffic source model* in order to simulate these sources. Using this model, any traffic source whose certain parameters (for example, average bit rate, burstiness factor, average number of cells in a burst etc.) are known, can be simulated. The simulations performed were *trace driven*, i.e., the traffic was first

generated and then the same traffic was applied to each of the three architectures.

5.2 Traffic Source Models

Traffic source characterization has been an extensive area of research [29]. The essential characteristics of a particular traffic source are all those parameters that are needed to completely characterize the randomness in the source. Such characteristics are important during the negotiation of required quality of service (QoS) at call setup. The essential traffic parameters of a source are used to develop a traffic model for the source. Traffic source models are important not just during the negotiation of QoS, but are also important for admission control, traffic policing, and during traffic control. They are also needed to predict the generation times of cells during the simulation of that source [29].

During the lifetime of a virtual connection, the corresponding traffic source will be in one of two states, *active* or *idle*. During the active state, the source is transmitting cells at some given rate. Depending on the type of source, each active state may be followed by an idle period during which the source is silent. This model is known as the *ON-OFF model*. There is a general belief that, with the exception of CBR sources, all other traffic sources exhibit this cyclic behavior (see Figure 5.1). For CBR-sources, there is no idle period.

There are other source models as well, such as the *generally modulated Deter-*

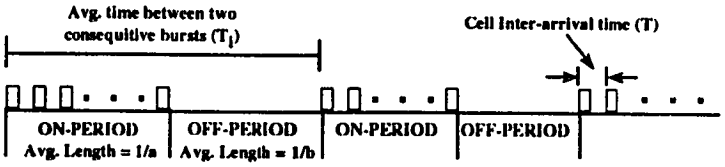

Figure 5.1: On-Off source model.

ministic Process Model (GMDP) or the *Markov Modulated Poisson Process Model (MMPP)* [29]. However, the ON-OFF model is the least complex and is the most widely used by researchers to model ATM traffic sources. Furthermore, this basic ON-OFF model is flexible enough to accommodate most of the existing traffic sources with a reasonable accuracy.

The cells generated during the same ON-period form a *burst*. Furthermore, it is always assumed that successive active and idle periods are statistically independent. As suggested by ITU-T, the length of the active period as well as that of the idle period are exponentially distributed, with average lengths $\frac{1}{a}$ and $\frac{1}{b}$ respectively (see Figure 5.1).

Several parameters have been identified, which together, completely characterize an ON-OFF traffic source. These are,

T: **Cell interarrival time.** This is the time between the arrival of the first bit of a cell and the first bit of the next consecutive cell from a given source

a^{-1}: **Mean value of active period.**

b^{-1}: **Mean value of idle period.**

p: **Peak cell arrival rate.** This is the cell arrival rate when the source is in the ON state that is $p = 1/T$, where T is the time between two consecutive cell arrivals during the ON period (see Figure 5.1).

m: **Average cell arrival rate.** This is the cell arrival rate over the entire lifetime of the connection of the source that is

$$m = p \times \frac{a^{-1}}{a^{-1} + b^{-1}}$$

β: **Traffic burstiness.** A large value for this parameter indicates a very bursty source. The burstiness is

$$\beta = \frac{p}{m}$$

t_{on}: **Average duration of active state.** This average is computed over the entire lifetime of the connection and is same as a^{-1}

Equivalently, a traffic source can be characterized by the following three traffic parameters [29]:

R_p: **Peak cell arrival rate.** $R_p = p = 1/T$.

B: **Average number of cells in a burst.** $B = a^{-1}/T$.

T_i: **Average time between two consecutive bursts.** $T_i = a^{-1} + b^{-1}$.

Typical values [29, 30] for the traffic parameters for the various traffic source types are summarized in Table 5.1.

Type of Source	B in cells	Average bit rate $m \times 384$ *bps*	Burstiness β	Cell Loss Tolerance
CBR	N/A	64 Kbps	1	10^{-4} to 10^{-6}
Connectionless data	200	700 Kbps	as high as 1000	10^{-12}
Connection oriented data	200	25 Mbps	as high as 1000	10^{-12}
VBR video	2	25 Mbps	2 to 5	10^{-10}
Background data/video	3	1 Mbps	2 to 5	10^{-9} to 10^{-10}
VBR video/data	30	21 Mbps	2 to 5	10^{-9}
Slow video	3	6 Mbps	2 to 5	10^{-12}

Table 5.1: Gains of cells.

In our simulation study, we assume that the following three parameters are known about each source: (1) the average bit rate, (2) the mean burst length B in cells, and (3) the burstiness factor β. As recommended by ITU-T, we assume also that the active and idle periods are exponentially distributed with parameters a and b respectively. The parameters a, b, as well as other parameters (such as T, and p) can all be calculated from the knowledge of m, B, and β. Next, we illustrate this calculation with an example.

Example:

Assume that a VBR Video/Data source has an average bit rate of 21 $Mbps$, an average number of cells in a burst $B = 30$, and a burstiness $\beta = 5$. Then the peak bit rate is $5 \times 21Mbs = 105Mbps$. The payload of each cell is 384 bits. Hence, $T = (384/105) \cdot 10^{-6} = 3.657\mu s$. Now, knowing the value of B and T and using the relation $B = a^{-1}/T$, one can readily obtain the value of a^{-1}. For this example,

$a^{-1} = B \cdot T = 109.71\mu sec$. Finally, using the relation $\beta = (a^{-1} + b^{-1})/a^{-1}$, one can also find that $b^{-1} = (\beta - 1)a^{-1} = 438.84\mu sec$.

Once the parameters a^{-1}, b^{-1} and T have been calculated, the cell generation can be simulated according to the ON-OFF model with exponentially distributed active and idle periods. We know that the exponential distribution can be generated from uniform distribution as follows.

Let y be an exponentially distributed random variable with mean equal to $1/\lambda$, $F(y)$ be the cumulative distribution function of y. Then,

$$F(y) = 1 - e^{-y \cdot \lambda} \tag{5.1}$$

Let,

$$z = F(y)$$

It can be proved that z is uniformly distributed over the interval (0, 1). Since $z = F(y)$,

$$z = 1 - e^{-y \cdot \lambda} \tag{5.2}$$

$$\Longrightarrow y = -(1/\lambda) \cdot ln\ (1 - z) \tag{5.3}$$

Since z and $(1 - z)$ have the same distribution,

$$y = -(1/\lambda) \cdot ln\ z \tag{5.4}$$

By taking $(1/\lambda)$ to be a^{-1} and b^{-1}, we can generate exponentially distributed durations for the active and idle periods respectively, for each source in a given time slot.

Using these values, we can calculate the number of cells in a given active period (i), using the equation, $B_i = a_i^{-1}/T$.

5.3 Simulation Results

In this section, we study the cell loss characteristics of the TBSF, PBSF and PTBSF switches when subjected to a variety of ATM traffic mixes. We implemented a program, which simulates any of the three switches. The simulator takes as input the switch type (TBSF, PBSF, or PTBSF), the number of inputs (N), and the number and types of each traffic source. The simulator has the capability of generating traffic according to any of the sources listed in Table 5.1. Following CCITT recommendation, all VBR sources follow the aforementioned *ON-OFF* model, where each VBR source is characterized by three parameters: its average cell rate m, the average number of cells per burst B, and the burstiness factor β.

It is impractical to simulate the three Banyan switches for all possible traffic mixes. Instead, we limited ourselves to some of the typical workloads expected by an ATM switch. We experimented with the following typical traffic mixes.

Traffic Mix 1: The sources consist of 20% Video, 50% Voice, and 30% Data. The destinations are selected with equal probabilities.

Traffic Mix 2: 40% of the sources are VBR Data/Video, 20% Voice, 20% Connectionless Data, and 20% Connection-Oriented Data. The destinations are se-

lected with equal probabilities.

Traffic Mix 3: The sources are as in *Traffic Mix 2*, but with output concentration, where only odd (respectively even) numbered output ports are selected.

Traffic Mix 4: The sources are as in *Traffic Mix 3*. Here also we perform output concentration, but this time only either the upper (respectively lower) output ports are selected.

Traffic-2 subjects the switch to a much larger workload than that generated by *Traffic-1*. The reason is that the second traffic mix has a higher percentage of burstier sources with larger bandwidth requirements. The change in performance behavior from *Traffic-1* to *Traffic-2* will expose the sensitivity of the switch to the workload. *Traffic-3* and *Traffic-4* generate as high a workload as *Traffic-2*, with the important difference that the level of internal (internal links) as well as external (output ports) contention is increased.

We simulated the three switch architectures under the above traffic mixes. Below, we summarize our findings with the objective of getting answers to the following questions:

1. How does the parallel tree architecture compare with the tandem and piled topologies under various traffic characteristics?

2. What is the effect of workload on cell loss, that is, how sensitive is the performance of the switches when the traffic mix is changed?

3. What is the effect of N on cell loss, that is, will there be any noticeable cell loss degradation when the number of inputs is increased?

Comparison of TBSF, PBSF, and PTBSF

Figures 5.2 and 5.6 show the variation of cell loss rate as a function of the number of SW-Banyans for TBSF, PBSF, and PTBSF. The figures correspond to N=32 and N=64 respectively, and a workload of type *Traffic-1*. PTBSF exhibited superior performance (lower loss rates) than the other two switches. PBSF exhibited the poorest behavior, with the cell loss rate saturating at about 10^{-6} or above from level 3 and onward. Such cell loss rate is not tolerated by most traffic sources. Hence PBSF is not a suitable architecture for an ATM switch fabric. Figures 5.2 and 5.6 illustrate another important fact. When the switch size is doubled from N=32 to N=64, we observed that the cell loss rate almost doubled for TBSF. On the other hand, both PBSF and PTBSF were not as sensitive to changes in N, and exhibited almost the same cell loss rates for both values of N. This observation requires further experimentation[1].

[1]We could not simulate the three switches for larger values of N because of excessively large run times

Effect of the workload on cell loss in TBSF, PBSF, and PTBSF

Figures 5.6-5.9 show the cell loss performance of TBSF, PBSF, and PTBSF for N=64 and for *Traffic-1*, *Traffic-2*, *Traffic-3*, and *Traffic-4* respectively. We observe that among the three switches, TBSF is the least sensitive to a change in the traffic mix. Both PBSF and PTBSF are sensitive to a change in the workload. However, we noticed that the sensitivity of PTBSF to the workload becomes less and less noticeable as we increase the number of levels (the number of SW-Banyans).

Figures 5.8 and 5.9 correspond to the traffic mix with odd/even and upper/lower output concentration respectively. Both traffic mixes generate more internal collision than traffic with no output concentration. The reader should be able to observe that for all three switches the cell loss performance for the lower/upper output concentration is noticeably worse than the case of odd/even output concentration. The reason is that lower/upper output concentration generates much more internal collisions than the odd/even output concentration. The reader should also note that PTBSF still exhibits superior performance than TBSF or PBSF for these traffic mixes.

All figures show the cell loss rate as a function of the number of SW-Banyans varying from 1 to 5. With the PBSF switch, we always reach saturation for small values of the number of SW-Banyans (≤ 3). For the PTBSF switch, the simulator

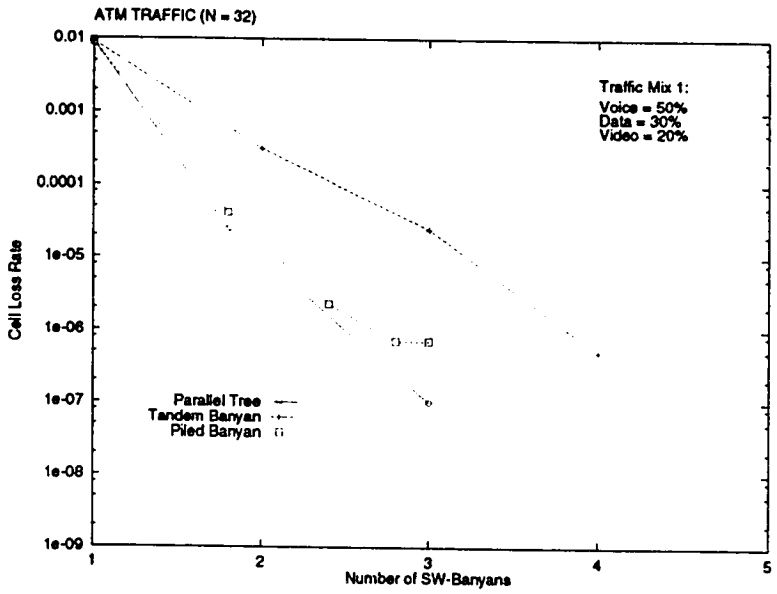

Figure 5.2: Cell loss rate versus the number of SW-Banyans for N=32 inputs, with 50% voice sources, 30% data, and 20% video. The destinations have equal probabilities of being selected

indicates a cell loss rate equal to zero when the number of SW-Banyans is greater than or equal to 5. For a TBSF switch with 5 levels, the cell loss rate varied between $5 \cdot 10^{-6}$ (*Traffic-3/4*) and zero (*Traffic-1/2*).

5.4 Conclusion

In this chapter we presented the performance evaluation of the three architectures, PTBSF, TBSF and PBSF under genuine ATM traffic conditions. The ATM traffic sources were modeled using the ON-OFF model. The ON-OFF model assumes that the sources are in just two states, either transmitting or idle. Any source whose certain parameters (B, m and β) are known can be simulated using this model.

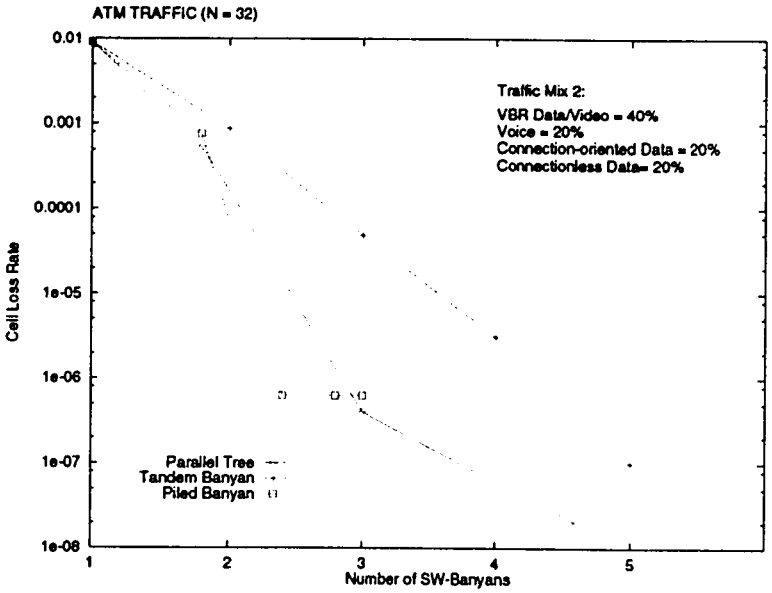

Figure 5.3: Cell loss rate versus the number of SW-Banyans for N=32 inputs, with 20% voice sources, 20% connection-oriented data, and 20% connectionless data and 40% VBR Video/Data. The destinations have equal probabilities of being selected.

The three architectures were subjected to various realistic traffic workloads generated by the traffic source simulator. The interarrival times of the cells generated by VBR sources are usually exponentially distributed. Hence, the traffic can vary considerably from one run (of the simulator) to the other. Thus, we adopted a *trace driven* methodology for the performance evaluation of the architectures under test, i.e., exactly the same traffic was applied to all the architectures. This helped us in making an accurate evaluation of the architectures under test. Based on the results, PTBSF exhibited excellent performance under all studied test conditions. The cell loss rate of PBSF saturates again even though the load is below 100% for ATM traffic conditions under which we performed our simulations.

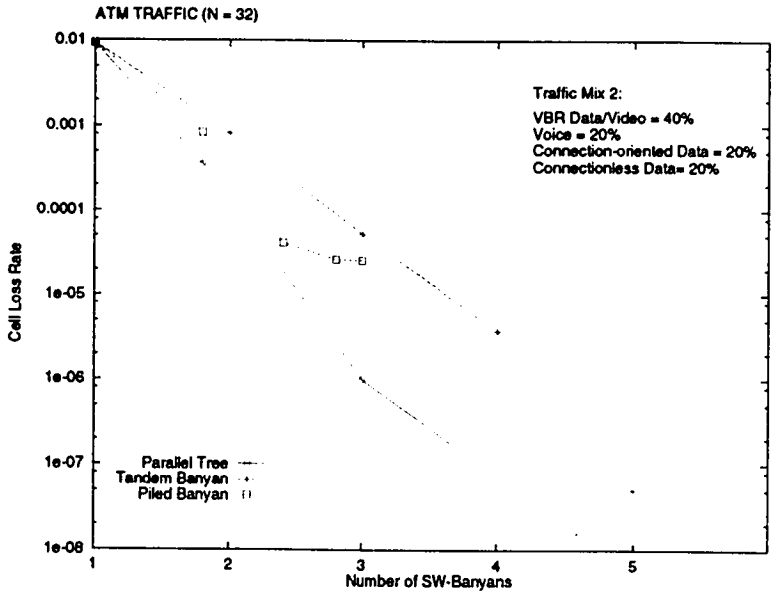

Figure 5.4: Cell loss rate versus number of SW-Banyans for N=32 inputs, with 20% voice sources, 20% connection-oriented data, 20% connectionless data and 40% VBR Video/Data. Output concentration, with either odd or even numbered destinations selected.

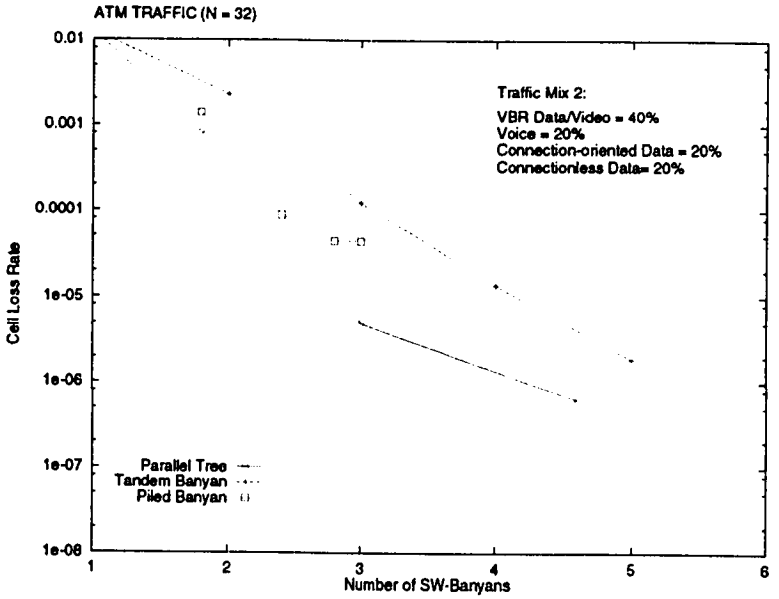

Figure 5.5: Cell loss rate versus number of SW-Banyans for N=32 inputs, with 20% voice sources, 20% connection-oriented data, and 20% connectionless data and 40% VBR Video/Data. Output concentration, with either upper or lower half of the destinations selected.

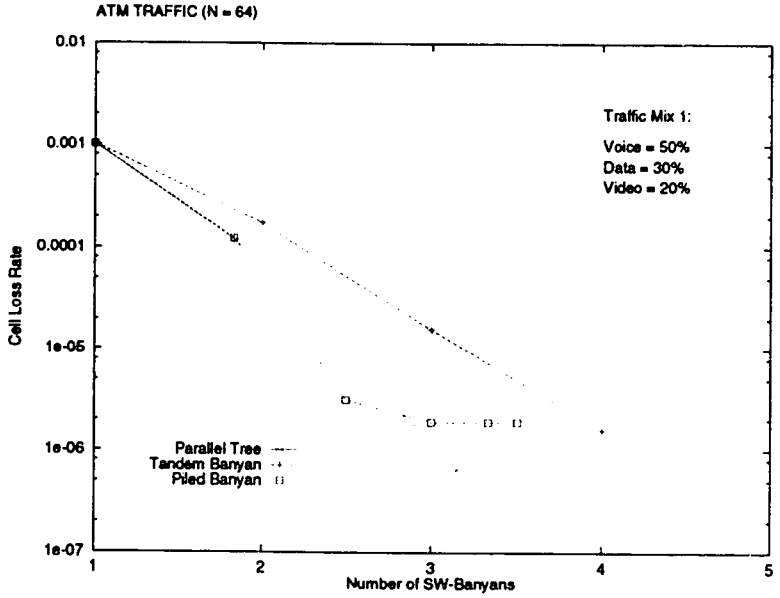

Figure 5.6: Cell loss rate versus the number of SW-Banyans for N=64 inputs, with 50% voice sources, 30% data, and 20% video. The destinations have equal probabilities of being selected

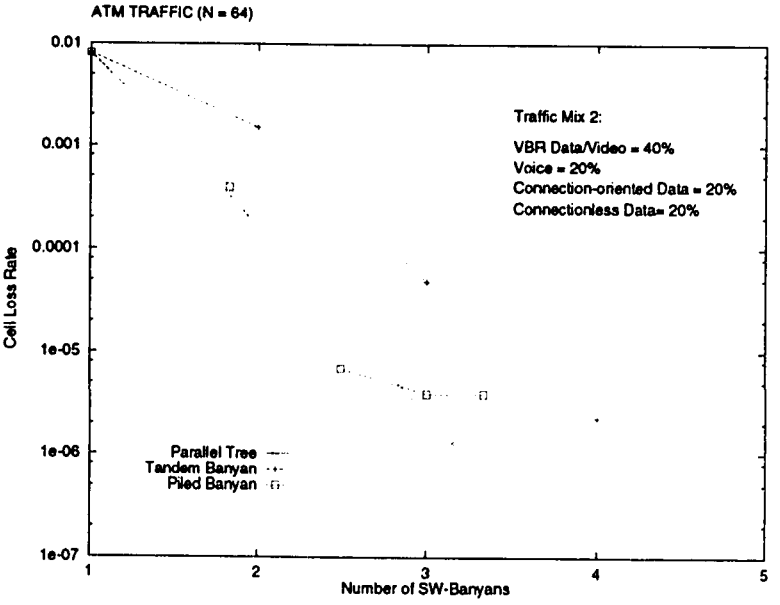

Figure 5.7: Cell loss rate versus the number of SW-Banyans for N=64 inputs, with 20% voice sources, 20% connection-oriented data, and 20% connectionless data and 40% VBR Video/Data. The destinations have equal probabilities of being selected.

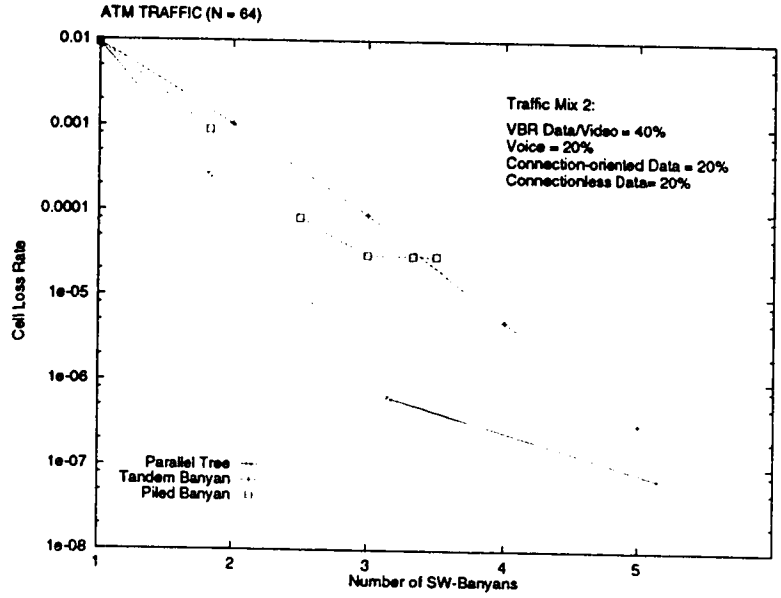

Figure 5.8: Cell loss rate versus number of SW-Banyans for N=64 inputs, with 20% voice sources, 20% connection-oriented data, 20% connectionless data and 40% VBR Video/Data. Output concentration, with either odd or even numbered destinations selected.

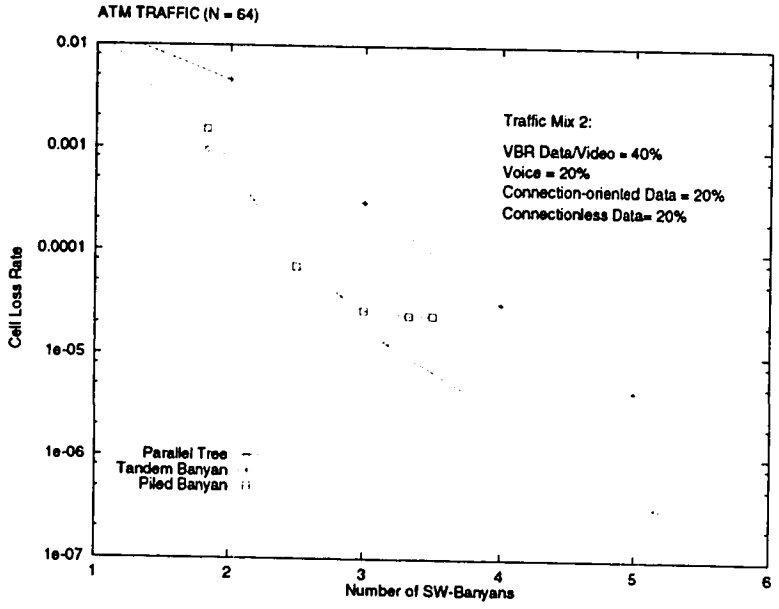

Figure 5.9: Cell loss rate versus number of SW-Banyans for N=64 inputs, with 20% voice sources, 20% connection-oriented data, 20% connectionless data and 40% VBR Video/Data. Output concentration, with either upper or lower half of the destinations selected.

Chapter 6

Conclusion

6.1 Summary

Significant advances in fiber optics technology have made available enormous amount of transmission bandwidth. This has provided a large impetus to the emergence of new applications which require much higher bandwidth than is possible with the current networking technology. Moreover, with the maturing of the VLSI technology, the device dimensions have shrunk considerably, so that it is now possible to put several millions of transistors on a single chip! Both higher speeds and higher levels of circuit integration are now possible [31].

Thus, the challenge facing the telecommunications community today, is to implement a network which can support any kind of service, existing as well as emerging, in an integrated fashion and which will also have broadband capabilities [32], [31],

[33], [34]. This is the objective behind implementing the *Broadband Integrated Services Digital Network* (B-ISDN).

The various services which are intended to be supported through B-ISDN may have diverse traffic characteristics in terms of bandwidth, delay, and packet loss requirements. Moreover, multimedia applications consisting of a combination of voice, data, video and image simultaneously, may have varied service requirements within the same application. This adds to the complexity of designing such a network which can support this wide variety of applications in an integrated, efficient and cost-effective manner. Before B-ISDN becomes a reality, three main issues must be resolved [35]:

- The development of network protocols which can support traffic sources with diverse service requirements.
- The design of switches which can provide switching of packets at a very high speed, have high throughput and low latency and are capable of being implemented with the current VLSI technology.
- Development of a control strategy which will guarantee *quality of service* (QoS) to every user, irrespective of the network traffic conditions.

The third is an open issue and not even a well defined problem [28].

Asynchronous Transfer Mode has been chosen as the switching and multiplexing technology for B-ISDN because it can support traffic sources with diverse require-

ments, including multi-media applications in an efficient and cost-effective manner. Much standardization activity is going on with respect to ATM [7], [32], [36].

ATM is a switch based technology. Thus, central to its success is the development of a switch architecture which can perform routing of cells at a very high speed with low loss and latency. Great research efforts are on in this regard and many switch architectures have been proposed recently [5], [6], [11], [12], [13], [14], [37].

In this thesis, we made an endeavor to understand the ATM technique and the various open problems. We focussed on the switching fabric for ATM networks. Our goal was to design a switch architecture which suits ATM requirements and overcomes many of the drawbacks found in the current architectures. Thus, we had the following design objectives in our mind at the inception of this research project:

1. The switch should be able to route cells '*on the fly*' with no buffering nor re-circulation to minimize processing time (switching delay) and hardware complexity.

2. The switch should guarantee acceptable cell loss rates with reasonable amount of hardware resources.

3. The *delay variance* of the cells over a given line, which is important, especially for real time applications, should be low.

4. The architecture should be scalable in the sense that increasing switch resource should always be rewarded by noticeable increase of the throughput.

We set about our task by making an in-depth study of the various aspects of the ATM technique, its cell structure, switching, protocol reference model, basic ATM switch architecture etc. Chapter 1 describes much of the effort in this direction.

In Chapter 2, we made a comprehensive literature survey of the various existing switch architectures. We presented a classification of the existing switch architectures based on the physical connection between the input and output ports of the switch fabric. Based on this classification, most of the existing ATM switch architectures may be broadly classified into time and space division architectures. We discussed both types of architectures along with some of their prototypes. The trend is towards space division architectures because they can support multiple connections simultaneously, and thus can provide high bandwidth. Among the space division architectures, banyan networks are the most popular for building ATM switches because of their numerous desirable characteristics. We discussed several architectures based on banyan networks along with their merits.

In Chapter 3, we presented a new space division switch architecture based on banyan networks called the *Parallel-Tree Banyan Switching Fabric* (PTBSF). PTBSF is based on a 3-D arrangement of banyans in a tree structure. The routing algorithm followed by the switch ensures reasonably uniform distribution of cells over the entire switching fabric, thus minimizing cell loss (as seen from the performance evaluation done in Chapters 4 and 5). Hardware resource requirements of the switch were also discussed. The architecture was contrasted with two other well

known architectures, the tandem banyan (TBSF)and the piled banyan (PBSF).

Chapter 4 presented the performance evaluation of the three architectures, PTBSF, TBSF and PBSF under independent uniform traffic pattern. Though uniform traffic does not represent actual ATM traffic sources, it is used as a common platform in order to compare the performance of different architectures because of its simplicity. We built analytical and simulation models for the above three architectures.

For the simulation model under uniform traffic conditions, we considered three cases: 1) cells with destinations selected randomly, 2) cells with destinations selected according to some fixed pattern and, 3) cells whose destinations are a perfect permutation of the set of output ports. PTBSF performed reasonably well for the first two cases compared to the other two architectures, especially for smaller switch sizes ($N \leq 128$). Performance evaluation under the third case, i.e., permutation traffic, suggests that TBSF is better suited to handle this kind of traffic[1]. PTBSF is superior to TBSF under this kind of traffic only in case of small switch sizes ($N = 32$ and below). The throughput of PBSF saturates after reaching a certain value under all three cases of uniform traffic, no matter how many levels are added.

We also evaluated the performance of the three architectures PTBSF, TBSF and PBSF under realistic ATM traffic workload. The results of this assessment were described in Chapter 5. The traffic sources which are expected in real ATM

[1]Note that permutation traffic is generally encountered only in synchronous transfer mode networks.

networks were modeled using the ON-OFF model. We built an ATM traffic source simulator based on this model. The traffic generated by these simulated traffic sources was applied to the simulators of the three switch architectures to evaluate the performance of each. The results show the superiority of PTBSF compared to the other two architectures, especially for small switch sizes under all studied conditions.

Thus, we were largely successful in our objectives of designing a switch architecture which could route cells on the fly, have a high throughput and low latency and whose performance is scalable. A thorough performance evaluation under varied traffic conditions vindicates this claim.

To conclude, we can say that although, a 3-D arrangement of banyan networks may result in more hardware resources in terms of interconnection wires between the planes, it certainly is a better approach in terms of average latency and delay jitter. Delay jitter is critical, especially for real time applications. Moreover, a high throughput may be achieved if proper routing of cells is performed within the switching fabric.

6.2 Future Work

We demonstrated through our work, the feasibility and advantages of a 3-D arrangement of banyan networks for building ATM switches. The future scope of our work

involves exploring the following aspects.

- Alternative *load balancing* techniques in order to distribute the traffic more uniformly over the tree structure may be explored instead of the *odd-even* strategy that we employed in the case when two conflicting cell arrive at a switching element.

- Other parallel architectures, based on a 3-D arrangement of banyan networks may be explored. For example, another parallel architecture may be a hybrid of tandem and PTBSF.

- Other architectures with less interconnection requirements may be investigated.

- We concentrated only on the switching fabric. Other aspects of an ATM switch may be taken into consideration, like, multiplexing at the input and de-multiplexing at the output.

Bibliography

[1] Rainer Handel and Manfred N. Huber. *Integrated Broadband Networks - An Introduction to ATM-Based Networks.* Addison-Wesley, Reading, MA, 1991.

[2] Ronald J. Vetter. Atm concepts, architectures and protocols. *Communications of the ACM*, 38(2): pp. 31–38, Feb 1995.

[3] Reza Rooholamini, Vladimir Cherkassky, and Mark Garver. Finding the right ATM switch for the market. *Computer*, : pp. 17–28, Apr. 1994.

[4] Fouad A. Tobagi. Fast packet switch architectures for broadband integrated services digital network. *Proceedings of the IEEE*, 78(1): pp. 133–167, Jan. 1990.

[5] Fouad A. Tobagi, Timothy Kwok, and Fabio M. Chiussi. Architecture, performance, and implementation of the tandem banyan fast packet switch. *IEEE J. Selected Areas in Communications*, 9(8): pp. 1173–1193, Oct. 1991.

[6] Toshihiro Hanawa et al. Multistage Interconnection Networks with multiple outlets. *1994 International Conference on Parallel Processing*, :I-1 - I-8, 1994.

[7] Martin de Prycker. *Asynchronous Transfer Mode - solution for broadband ISDN.* Ellis Horwood, 1991.

[8] M. Devault, J. Cochennec, and M. Servel. The Prelude ATD experiment: assessments and future prospects. *IEEE J. Selected Areas in Communications*, 6(9): pp. 1528–1537, Dec. 1988.

[9] H. Suzuki et al. Output-buffer switch architecture for asynchronous transfer mode. *Proc. Int. Conf. on Communications, Boston, MA*, : pp. 4.1.1–4.1.5, June 1989.

[10] Y. Yeh, M. Hluchyj, and A. Acampora. The Knockout Switch: A simple, modular architecture for high-performance packet-switching. *IEEE J. Selected Areas in Communications*, SAC-5(8): pp. 1274–1283, Oct. 1987.

[11] M. J. Narasimha. The Batcher-banyan self routing network: universality and simplification. *IEEE Trans. Commun.*, 36(10): pp. 1175–1178, Oct 1988.

[12] K. E. Batcher. Sorting networks and their applications. *AFIPS Proc. 1968 Spring Joint Computer Conf.*, 32: pp. 307–314, 1968.

[13] A. Huang and S. Knauer. Starlite: A wideband digital switch. *Proc. GLOBECOM 84, Atlanta, GA*, : pp. 121–125, 1984.

[14] J. Giacopelli et al. Sunshine: A high performance self-routing broadband packet-switch architecture. *IEEE J. Selected Areas in Communications*, : pp. 1289–1298, Oct 1991.

[15] M. Kawarasaki and B. Jabbari. B-ISDN architecture and protocol. *IEEE J. Selected Areas in Communications*, 9(9): pp. 1405–1415, Dec. 1991.

[16] K. Chipman et al. Medical applications in a B-ISDN field trial. *IEEE Journal of Selected Areas in Commun.*, 10(7): pp. 1173–1187, Sept. 1992.

[17] D. Delisle and L. Pelamourgues. B-ISDN and how it works. *IEEE Spectrum*, 28(8): pp. 39–42, Aug. 1991.

[18] M. DePrycker, R. Peschi, and Landegem. B-ISDN and the OSI protocol reference model. *IEEE Network*, 7(2): pp. 10–18, Mar. 1993.

[19] T. Suzuki. Atm adaptation layer protocol. *IEEE Commun. Mag.*, 32(4): pp. 80–83, Apr. 1994.

[20] P. Newman. ATM technology for corporate networks. *IEEE Commun. Magazine*, 30(4): pp. 90–101, Oct 1992.

[21] G. Hayward and et al. CMOS VLSI applications in broadband circuit switching. *IEEE Journal of Selected Areas in Commun.*, SAC-5(8): pp. 1231–1241, Oct. 1987.

[22] H. Ahmadi and W. Denzel. A survey of modern high-performance switching techniques. *IEEE J. Selected Areas in Communications*, 7(7): pp. 1091–1103, Sept. 1989.

[23] J. Y. Hui and E. Arthurs. A broadband packet switch for integrated transport. *IEEE J. Selected Areas in Communications*, 5(8): pp. 1264–1273, Oct. 1987.

[24] C. T. Lea. Multi-Log_2N networks and their applications in high speed elctronic and photonic switching systems. *IEEE Trans. Commun.*, 38(10): pp. 1740–1749, Oct. 1990.

[25] C. P. Kruskal and M. Snir. The performance of multistage interconnection networks for multiprocessors. *IEEE Trans. Comput.*, C-32(12): pp. 1091–1098, Dec 1983.

[26] M. Kumar and J. R. Jump. Performance of unbuffered shuffle exchange networks. *IEEE Trans. Comput.*, C-35(6):573–577, Jun. 1986.

[27] R. Y. Awdeh and H. Mouftah. The Expanded Delta Fast Packet Switch. *Proc. IEEE SUPERCOMM/ICC'94, New Orleans, LA*, : pp. 397–401, May 1994.

[28] Ra'ed Y. Awdeh and H. T. Mouftah. Survey of ATM switch architectures. *Computer Networks and ISDN Systems*, 27: pp. 1567–1613, Nov. 1995.

[29] G. D. Stamoulis, M. E. Anagnostou, and A. D. Georgantas. Traffic source models for ATM networks: a survey. *Computer Communications, Butterworth-Heinemann Ltd*, 17(6): pp. 428–438, June 1994.

[30] Ken Dubose and Hyong S. Kim. An Effective Bit Rate/Table Lookup Based Admission Control Algorithm for the ATM B-ISDN. *Proceedings of the* 17^{th} *Conference on Local Computer Networks, Minneapolis*, : pp. 20–29, Sept. 1992.

[31] J. E. Berthold. High speed integrated electronics for communications systems. *Proc. IEEE*, 78(3): pp. 486–511, Mar. 1990.

[32] J. J. Bae and T. Suda. Survey of traffic control schemes and protocols in ATM networks. *Proc. IEEE*, 79(2): pp. 170–189, Feb. 1991.

[33] C. S. Cooper. High-speed networks: the emergence of technologies for multi-service support. *Computer Commun.*, 14(1): pp. 27–43, Feb. 1991.

[34] M. De Prycker. Evolution from ISDN to B-ISDN: a logical step towards ATM. *Computer Commun.*, 12(3): pp. 141–146, June 1989.

[35] A. Pattavina. Non-blocking architectures for ATM switching. *IEEE Commun. Mag.*, (2): pp. 38–48, Feb. 1993.

[36] J. Y. L. Boudec. The Asynchronous Transfer Mode: a tutorial. *Computer Networks and ISDN Systems*, 24: pp. 279–309, 1992.

[37] P. C. Wong and M. S. Yeung. Design and Analysis of a Novel Fast Packet Switch–Pipeline Banyan. *IEEE/ACM Transactions on Networking*, 3(1): pp. 63–69, Feb. 1995.

Vitae

- Wasif Hasan
- Born in 1972 at Aligarh, India
- Received Bachelor of Science (Engineering) (**B.S (Engg)**) degree in Electronics and Communication Engineering from Aligarh Muslim University, Aligarh, India in 1993
- Joined the Department of Computer Engineering at King Fahd University of Petroleum and Minerals (**KFUPM**), Dhahran, Saudi Arabia as a Research/Teaching Assistant in January 1994
- Received Master of Science (**M.S.**) degree in Computer Engineering from KFUPM, Saudi Arabia in 1996